

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:38:49 ; Search time 56 Seconds
(without alignments)
4432.963 Million cell updates/sec

Title: US-09-734-672-4

Perfect score: 9649

Sequence: 1 MDLSALRVEYQNVINAMQK.....LYQCQLDTYLIPQIPHSHY 1863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

```
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9649	100.0	1863	17 AAR81535	BRCA1 mutant from
2	9649	100.0	1863	19 AAW76099	Human BRCA1 (Omi2 p
3	9649	100.0	1863	21 AAB24219	Human BRCA1 (Omi3)
4	9642	99.9	1863	17 AAR97128	BRCA1, breast and
5	9642	99.9	1863	17 AAR81481	BRCA1, Homo sapie
6	9642	99.9	1863	17 AAR91208	BRCA1, breast and
7	9642	99.9	1863	20 AAY32033	Human BRCA1 protei
8	9642	99.9	2353	22 ABG01696	Novel human diagno
9	9639	99.9	1863	17 AAR81490	BRCA1 mutant from
10	9639	99.9	1863	17 AAR81500	BRCA1 mutant from

11	9639	99.9	1863	17 AAR81522	BRCA1 mutant from
12	9639	99.9	1863	17 AAR81536	BRCA1 mutant from
13	9639	99.9	1863	17 AAR81540	BRCA1 mutant from
14	9638	99.9	1863	17 AAR81529	BRCA1 mutant from
15	9638	99.9	1863	17 AAR81532	BRCA1 mutant from
16	9638	99.9	1863	17 AAR81533	BRCA1 mutant from
17	9638	99.9	1863	17 AAR81534	BRCA1 mutant from
18	9638	99.9	1863	17 AAR81514	BRCA1 mutant from
19	9638	99.9	1863	17 AAR81543	BRCA1 mutant from
20	9638	99.9	1863	17 AAR81545	BRCA1 mutant from
21	9637	99.9	1863	17 AAR81485	BRCA1 mutant from
22	9637	99.9	1863	17 AAR81493	BRCA1 mutant from
23	9637	99.9	1863	17 AAR81505	BRCA1 mutant from
24	9637	99.9	1863	17 AAR81511	BRCA1 mutant from
25	9637	99.9	1863	17 AAR81538	BRCA1 mutant from
26	9636	99.9	1863	17 AAR81486	BRCA1 mutant from
27	9636	99.9	1863	17 AAR81509	BRCA1 mutant from
28	9636	99.9	1863	17 AAR81537	BRCA1 mutant from
29	9636	99.9	1863	17 AAR81542	BRCA1 mutant from
30	9636	99.9	1863	17 AAR81544	BRCA1 mutant from
31	9635	99.9	1863	17 AAR81524	BRCA1 mutant from
32	9635	99.9	1863	17 AAR81546	BRCA1 mutant from
33	9635	99.9	1863	18 AAW25522	Human BRCA1 consen
34	9635	99.9	1863	19 AAW79665	BRCA1 (Omi1) prote
35	9635	99.9	1863	19 AAW76100	Human BRCA1 (Omi3 p
36	9635	99.9	1863	19 AAW76098	Human BRCA1 (Omi1 p
37	9635	99.9	1863	21 AAB24217	Human BRCA1 (Omi1)
38	9635	99.9	1863	21 AAB24218	Human BRCA1 (Omi2)
39	9634	99.8	1863	17 AAR81488	BRCA1 mutant from
40	9634	99.8	1863	17 AAR81519	BRCA1 mutant from
41	9634	99.8	1863	17 AAR81539	BRCA1 mutant from
42	9630	99.8	1863	17 AAR99440	BRCA1 allele #8403
43	9630	99.8	1863	17 AAR81497	BRCA1 mutant from
44	9630	99.8	1863	18 AAW10003	Protein encoded by
45	9630	99.8	1863	19 AAW79889	Tumourigenic BRCA1

ALIGNMENTS

RESULT 1

AAR81535
ID AAR81535 standard; Protein; 1863 AA.

XX AC AAR81535;

XX DT 02-OCT-1996 (first entry)

XX DE BRCA1 mutant from PM06.

XX DE Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;

KW antibody production; germline alteration; probe; lesion neoplasia; human;

KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Misc-difference 871 /note= "p871L"

XX XX WO9605306-A2.

XX PD 22-FEB-1996.

XX PF 11-AUG-1995; 95WO-US10202.

XX PR 07-JUN-1995; 95US-0483553.

PR 12-AUG-1994; 94US-0289221.

PR 02-SEP-1994; 94US-0300266.

PR 16-SEP-1994; 94US-0308104.

PR 29-NOV-1994; 94US-0348824.

PR 24-MAR-1995; 95US-0409305.

PR 07-JUN-1995; 95US-0480784.

XX (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.
XX
XX Durocher F., Emi M., Nakamura Y., Simard J., Shattuck-Eidens DM;
DR WPI: 1996-139702/14.
DR N-PSDB; AAT17493.
XX
XX New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
XX
XX Claim 1: ; 218pp; English.
XX
XX AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
CC (see AAR81481 for wild type protein). These mutations can be used as
CC immunogens for antibody production. The mutant BRCA1 genes encoding
CC these sequences have at least 1 mutation or polymorphism in comparison
CC to the wild type cDNA (see AAT17438 for wild type). By detecting a
CC germline alteration in the wild type BRCA1 gene, a predisposition for
CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
CC isolated from a tissue sample from a subject has a probe, corresponding
CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an
CC allele-specific probe for a mutation of it), added to it. The conditions
CC allow for hybridisation of the probe to the mRNA, and any hybridisation
CC which occurs is detected. Alternatively the BRCA1 gene in the tissue
CC sample is isolated, and a shift in electrophoretic mobility of single
CC stranded DNA from the sample on a non-denaturing polyacrylamide gel
CC indicates a mutation. These methods of detection can also diagnose a
CC lesion neoplasia associated with the BRCA1 locus. The methods may be
CC used in gene therapy, protein replacement therapy and protein mimetics,
CC and may be used to screen for drugs in cancer therapy.
XX
SQ Sequence 1863 AA;

Query Match 100.0%; Score 9649; DB 17; Length 1863;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQVNIAMOKILECPICLELIKEPVSTKCDHIFCKFCLMLLNQKGPSP 60
DB 1 MDLSALRVEEVQVNIAMOKILECPICLELIKEPVSTKCDHIFCKFCLMLLNQKGPSP 60

QY 61 CPLCKNDITKRSLOSTRFPSQLVEELLKICAFQDITGLEYANSYNFAKKNNSPEHLKD 120
DB 61 CPLCKNDITKRSLOSTRFPSQLVEELLKICAFQDITGLEYANSYNFAKKNNSPEHLKD 120

QY 121 EYSIIQSMGYRNRARLLQSPENPSLQETSLVSQLSNIGTVRTLRTKORTPOKTSVYI 180
DB 121 EYSIIQSMGYRNRARLLQSPENPSLQETSLVSQLSNIGTVRTLRTKORTPOKTSVYI 180

QY 181 ELGSDSSEDVTNKAITYCSVGQDQLLQITPQGTREISLDSAKKAACEFSETDVTNTEHHQ 240
DB 181 ELGSDSSEDVTNKAITYCSVGQDQLLQITPQGTREISLDSAKKAACEFSETDVTNTEHHQ 240

QY 241 PSNNDLNTTEKRAARHPHEKPYQGSVSNLHVPECGTNTHASSLQHNSSLLLTKDRNVE 300
DB 241 PSNNDLNTTEKRAARHPHEKPYQGSVSNLHVPECGTNTHASSLQHNSSLLLTKDRNVE 300

QY 301 KAEFCNKSQPLGARSOHNRWAGSKETCNDRTPTSTKKVDLNDADPLCKERKNQKLPC 360
DB 301 KAEFCNKSQPLGARSOHNRWAGSKETCNDRTPTSTKKVDLNDADPLCKERKNQKLPC 360

QY 361 SENPRDTEDPWITLNSIQVNWFSRSDLLGSDSDSHGSESSEAKVADVLVDLNEVD 420
DB 361 SENPRDTEDPWITLNSIQVNWFSRSDLLGSDSDSHGSESSEAKVADVLVDLNEVD 420

QY 421 EYSGSSEKIDLLASDPHEALICKSERVHKSQSVESNIEDKIFGKTYRKASLPNLSHVTE 480
DB 421 EYSGSSEKIDLLASDPHEALICKSERVHKSQSVESNIEDKIFGKTYRKASLPNLSHVTE 480

QY 481 LIICAFVTEPQIIQERPLTNKLRKRPTSGLHPDEFIKKADLAVQKTPEMINQGTNOTE 540
DB 481 LIICAFVTEPQIIQERPLTNKLRKRPTSGLHPDEFIKKADLAVQKTPEMINQGTNOTE 540

QY 541 QNGOVMITNSGHNKTKGDSIQNEKNPNPIESLEKESAFKTKAEPPISSISNMLELENI 600
DB 541 QNGOVMITNSGHNKTKGDSIQNEKNPNPIESLEKESAFKTKAEPPISSISNMLELENI 600

QY 601 HNSKAPKNRLRRKSSRTHIALELVVSRNLSPNCTELQIDSCSSSEIEKKKKYQMPV 660
DB 601 HNSKAPKNRLRRKSSRTHIALELVVSRNLSPNCTELQIDSCSSSEIEKKKKYQMPV 660

QY 661 RHRNLQIMEGKEPATGAKKSNKNEQTSKRHSDTPPELKLITNAPGFTKCSNTSELKE 720
DB 661 RHRNLQIMEGKEPATGAKKSNKNEQTSKRHSDTPPELKLITNAPGFTKCSNTSELKE 720

QY 721 FVNPSPREEKEERLETVKVSNNNAEDPKDMLSGSERVLQTSRSESSISLVPGTDYGTQ 780
DB 721 FVNPSPREEKEERLETVKVSNNNAEDPKDMLSGSERVLQTSRSESSISLVPGTDYGTQ 780

QY 781 ESISLLEVLGKAKTBPNCVCVSOCAAFENPKGLIHGCKSKNNDTEGFKYPLGHEVNH 840
DB 781 ESISLLEVLGKAKTBPNCVCVSOCAAFENPKGLIHGCKSKNNDTEGFKYPLGHEVNH 840

QY 841 RETSIEMESELDAQYLQNTFKVSKROSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900
DB 841 RETSIEMESELDAQYLQNTFKVSKROSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900

QY 901 FECQKEENQKNEKNKPKVQTVNITAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRG 960
DB 901 FECQKEENQKNEKNKPKVQTVNITAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRG 960

QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTCKKNLLEENFEHSHSPREMGNEIP 1020
DB 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTCKKNLLEENFEHSHSPREMGNEIP 1020

QY 1021 STVSTISRNINRENVPKEASSNINEVGSSTNEVGSSTNEIGSDENIQAEIAGNRGPKL 1080
DB 1021 STVSTISRNINRENVPKEASSNINEVGSSTNEVGSSTNEIGSDENIQAEIAGNRGPKL 1080

QY 1081 NAMRLGLVQPEVYKQSLPGSNCKHPHKEIKOEYBEVQTVNTDFSPYLLISDNLEQPMGSS 1140
DB 1081 NAMRLGLVQPEVYKQSLPGSNCKHPHKEIKOEYBEVQTVNTDFSPYLLISDNLEQPMGSS 1140

QY 1141 HASQVCSSETPDDLLDGEIKEDETSFAENDIKESSAVFSKSVQKSELRSPPFTHTHLAQ 1200
DB 1141 HASQVCSSETPDDLLDGEIKEDETSFAENDIKESSAVFSKSVQKSELRSPPFTHTHLAQ 1200

QY 1201 GYRRGAKKLESSEENLSSEDEELPCFQHLILFGKVNNIPSQSTRHSTVATECLSKNTEENL 1260
DB 1201 GYRRGAKKLESSEENLSSEDEELPCFQHLILFGKVNNIPSQSTRHSTVATECLSKNTEENL 1260

QY 1261 LSLKNSLNDSCSNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
DB 1261 LSLKNSLNDSCSNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320

QY 1321 SKQMRHOSQSGVGLSDKELVSDDEERGTLGLENQNEEQSDNLSNLSGEAASCESETSVSE 1380
DB 1321 SKQMRHOSQSGVGLSDKELVSDDEERGTLGLENQNEEQSDNLSNLSGEAASCESETSVSE 1380

QY 1381 DCSGLSSOSDILTQQORDTMOHNLILKQOEAMAEVLEQHGCSQPSNSYPSIISDSSALE 1440
DB 1381 DCSGLSSOSDILTQQORDTMOHNLILKQOEAMAEVLEQHGCSQPSNSYPSIISDSSALE 1440

QY 1441 DLRNPEQSTSEKAVLTQSKSEYPIQNPESGLADKFEVSADSSTSKNKEPGRSSPSK 1500
DB 1441 DLRNPEQSTSEKAVLTQSKSEYPIQNPESGLADKFEVSADSSTSKNKEPGRSSPSK 1500

QY 1501 CPSLDDRWYHSCSGSLQNRNYPQSEELIKVVDVEEQLESGPHDLTETSYLPRQDLEG 1560
DB 1501 CPSLDDRWYHSCSGSLQNRNYPQSEELIKVVDVEEQLESGPHDLTETSYLPRQDLEG 1560

Qy 1561 TPYLESGISLSDPESDPSDRAPESARVGNIPSSTSALKVPQLKVAESQAASAAHTT 1620
Db 1561 TPYLESGISLSDPESDPSDRAPESARVGNIPSSTSALKVPQLKVAESQAASAAHTT 1620
Qy 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFARKHHTLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFARKHHTLTNLI 1680
Qy 1681 TEETHVVMKTDAEFVCERTLKYFLGAGGKWWVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDAEFVCERTLKYFLGAGGKWWVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
Qy 1741 VNGRHOGPKRARESQRKIFRGLIEICCYGFTTNMPTDQLEWMVOLCGASVVKLSFTL 1800
Db 1741 VNGRHOGPKRARESQRKIFRGLIEICCYGFTTNMPTDQLEWMVOLCGASVVKLSFTL 1800
Qy 1801 GTGVHPITVVQPDATWEDNGFHAIGOMCEAPVVTREWVLDSVALYQCOELDTYLIPQIPH 1860
Db 1801 GTGVHPITVVQPDATWEDNGFHAIGOMCEAPVVTREWVLDSVALYQCOELDTYLIPQIPH 1860
Qy 1861 SHY 1863
Db 1861 SHY 1863
RESULT 2
AAW76099
ID AAW76099 standard; Protein; 1863 AA.
XX
AC AAW76099;
XX
DT 18-NOV-1998 (first entry)
XX
DE Human BRCA1 omi2 protein.
XX
KW BRCA1; omi2; human; breast and ovarian cancer predisposing gene;
KW polymorphism; susceptibility; anti-oncogene; tumour suppressor;
XX chromosome 17q.
XX
OS Homo sapiens.
XX
PN US5750400-A.
XX
PD 12-MAY-1998.
XX
PF 12-FEB-1997; 97US-0798691.
XX
PR 12-FEB-1996; 96US-0598591.
PR 12-FEB-1997; 97US-0798691.
XX
PA (ONCO-) ONCORMED INC.
XX
PI Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;
PI Schelter DB, Zeng B;
XX
DR WPI; 1998-296774/26.
DR N-PSDB; AAV46449.
XX
PT BRCA1 omi gene coding sequences - useful for distinguishing between
PT polymorphisms and mutation(s) in the screening for disposition to
PT breast or ovarian cancer
XX
PS Disclosure; Column 47-56; 54pp; English.
XX
CC This sequence represents the human BRCA1 (breast and ovarian cancer
CC predisposing gene) omi2 gene. This sequence and polymorphic variations of
CC this sequence are useful for the identification of an individual who may
CC or may not have an increased susceptibility to breast or ovarian cancer.
CC The sequences used identify gene changes which are due to polymorphisms,
CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour
CC suppressor) which is involved in genetic inheritance of cancers,
CC especially breast and ovarian cancer. It is found at human chromosome 17q
CC which is known to be linked to cancer susceptibility, especially breast

CC cancer. Cells containing a mutation in this gene lose the wild-type
CC function of BRCA1 and are more susceptible to cancers.
XX
SQ Sequence 1863 AA;
Query Match 100.0%; Score 9649; DB 19; Length 1863;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MDLSALRVEEVQNVINAMOKILECPICLELILKEPVSFKCDHIFCKFCMLKLLNQKKGPSQ 60
Db 1 MDLSALRVEEVQNVINAMOKILECPICLELILKEPVSFKCDHIFCKFCMLKLLNQKKGPSQ 60
Qy 61 CPICKNDITKRSQESTRFSQVVEELKIIICAFOLDTGLEVANSYNPAKKENNSPEHLKD 120
Db 61 CPICKNDITKRSQESTRFSQVVEELKIIICAFOLDTGLEVANSYNPAKKENNSPEHLKD 120
Qy 121 EVSTIQSMGYRNRAKRLQLQSEPNPSLQETSLSVQLSNLGTVRTLRTKQRIQPKTSVYI 180
Db 121 EVSTIQSMGYRNRAKRLQLQSEPNPSLQETSLSVQLSNLGTVRTLRTKQRIQPKTSVYI 180
Qy 181 ELGSDSSEDTVNKATYCSVGDOELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDTVNKATYCSVGDOELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Qy 241 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGNTTHASSLQHENSLLLTKDRMNV 300
Db 241 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGNTTHASSLQHENSLLLTKDRMNV 300
Qy 301 KAFCNKSKQPLARSQHNWAGSKETCNDRRTPSTPEKKYVDLNADPLCERKENWKQLPC 360
Db 301 KAFCNKSKQPLARSQHNWAGSKETCNDRRTPSTPEKKYVDLNADPLCERKENWKQLPC 360
Qy 361 SENPRDTEVPWITLNSIOKVNEWFSRDELLGSDSDHGESNAKVADVDLNEVD 420
Db 361 SENPRDTEVPWITLNSIOKVNEWFSRDELLGSDSDHGESNAKVADVDLNEVD 420
Qy 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIDKIFGKTYRKASLPNLSHVTE 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIDKIFGKTYRKASLPNLSHVTE 480
Qy 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPPEMINQGTOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPPEMINQGTOTE 540
Qy 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPITSSISNMELELNI 600
Db 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPITSSISNMELELNI 600
Qy 601 HNSKAPKKNLRRKSTRHIALELVVSRNLSPNCTELOIDSCSSEETKKKYNOMPV 660
Db 601 HNSKAPKKNLRRKSTRHIALELVVSRNLSPNCTELOIDSCSSEETKKKYNOMPV 660
Qy 661 RHRNRLQMEGKEPATGAKSKNKPNEQTSKRHSDTDFPELKLTNAPGSFTKCSNTSELKE 720
Db 661 RHRNRLQMEGKEPATGAKSKNKPNEQTSKRHSDTDFPELKLTNAPGSFTKCSNTSELKE 720
Qy 721 FVNPSPREEKEKLETVKYSNNAEDPKDMLSGERVLOQTERSVESSISLVPGTDTGTQ 780
Db 721 FVNPSPREEKEKLETVKYSNNAEDPKDMLSGERVLOQTERSVESSISLVPGTDTGTQ 780
Qy 781 ESISLLEVTSLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRDNTGFKYPLGHEVNH 840
Db 781 ESISLLEVTSLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRDNTGFKYPLGHEVNH 840
Qy 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNGNAEEECATFFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNGNAEEECATFFSAHSGSLKKQSPKVT 900
Qy 901 FECEQKEENOGKNESTKPVQTVNITAGFPVVGQKDPVDNAKCSIKGGRFCLSSQFRG 960
Db 901 FECEQKEENOGKNESTKPVQTVNITAGFPVVGQKDPVDNAKCSIKGGRFCLSSQFRG 960

```
Qy 961 NETGLITPNKHGLQNPYRIPPIKSFVYKTKCKNLLLEENFEHSPEREMGNENIP 1020
Db 961 NETGLITPNKHGLQNPYRIPPIKSFVYKTKCKNLLLEENFEHSPEREMGNENIP 1020
Qy 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNGPKL 1080
Db 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNGPKL 1080
Qy 1081 NAMRLGVLPQPEYKOSLPGSNCKHPEIKKOEYEEVVQVNTDFSPYLLSDNLEOPMGSS 1140
Db 1081 NAMRLGVLPQPEYKOSLPGSNCKHPEIKKOEYEEVVQVNTDFSPYLLSDNLEOPMGSS 1140
Qy 1141 HASQVCSPTDLDLDDGEIKEDTSFAENDIKESSAVFSKVQKGLSRSPSPFTTHLAQ 1200
Db 1141 HASQVCSPTDLDLDDGEIKEDTSFAENDIKESSAVFSKVQKGLSRSPSPFTTHLAQ 1200
Qy 1201 GYRGAKKLESSEENLSSEDEELPCQHLHFGKVNIPISQSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRGAKKLESSEENLSSEDEELPCQHLHFGKVNIPISQSTRHSTVATECLSKNTEENL 1260
Qy 1261 LSLKNSLDCSNVILAKASOEHLSSEETKCSASLFSQSCSELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLDCSNVILAKASOEHLSSEETKCSASLFSQSCSELEDLTANTNTQDPFLIGS 1320
Qy 1321 SKQMRHQSQSGVGLSDKELVSDDEERGTLGLENNOEQSMDSNLGEASGCESETSVSE 1380
Db 1321 SKQMRHQSQSGVGLSDKELVSDDEERGTLGLENNOEQSMDSNLGEASGCESETSVSE 1380
Qy 1381 DCSGLSSQSDILITQORDTQWHLNKLQEMAELEAVLEQHSQSPNSYPSIISSSALE 1440
Db 1381 DCSGLSSQSDILITQORDTQWHLNKLQEMAELEAVLEQHSQSPNSYPSIISSSALE 1440
Qy 1441 DLNRPEOSTSEKAVLTQSQSSEYPISONPEGLSADKFEVSADSSTSKKNKEPVERSSPSK 1500
Db 1441 DLNRPEOSTSEKAVLTQSQSSEYPISONPEGLSADKFEVSADSSTSKKNKEPVERSSPSK 1500
Qy 1501 CPSLDDRWYMHSCGSLQNRNYPQBELIKVVDVEEQLEESGPHDLTETSYPRLQDLEG 1560
Db 1501 CPSLDDRWYMHSCGSLQNRNYPQBELIKVVDVEEQLEESGPHDLTETSYPRLQDLEG 1560
Qy 1561 TPYLESGISLFSDDPSDESDRAPESARVGNIPSTSTALKVPOLKVAESAQSPAAHHT 1620
Db 1561 TPYLESGISLFSDDPSDESDRAPESARVGNIPSTSTALKVPOLKVAESAQSPAAHHT 1620
Qy 1621 DTAGYNAMESVSREKPELTASTERNVKNRMSVMVSGLTPEEPFMYKFKARKHHTLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERNVKNRMSVMVSGLTPEEPFMYKFKARKHHTLTNLI 1680
Qy 1681 TEETHVVMKTDAEFVCERTLYFLGIAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDAEFVCERTLYFLGIAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
Qy 1741 VNGRNHQPGRKRESQDRKIFRGLETCCYGPPTNPTDQLEWVQLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQPGRKRESQDRKIFRGLETCCYGPPTNPTDQLEWVQLCGASVVKELSSFTL 1800
Qy 1801 GTGVHPITVVVQPDWNTDNGFHAIGQMCEAPVVTREWLDLSVALYQCOELDYILIPQIPH 1860
Db 1801 GTGVHPITVVVQPDWNTDNGFHAIGQMCEAPVVTREWLDLSVALYQCOELDYILIPQIPH 1860
Qy 1861 SHY 1863
Db 1861 SHY 1863

RESULT 3
ID AAB24219
XX AAB24219 standard; Protein; 1863 AA.
AC AAB24219;
XX
XX 07-FEB-2001 (first entry)
XX
```

```
DE Human BRCA1 (om13) protein sequence SEQ ID NO:6.
XX Human: BRCA1; chromosome 17; 17q21; breast cancer; ovarian cancer;
KW gene therapy; diagnosis; cytostatic; genetic susceptibility; mutation;
KW polymorphism; identification.
XX
OS Homo sapiens.
XX
PN US6130322-A.
PD 10-OCT-2000.
XX
XX 06-MAY-1998; 98US-0074476.
PR 12-FEB-1996; 96US-0598591.
PR 12-DEC-1997; 97US-0798691.
XX
XX (GENE-) GENE LOGIC INC.
PI Zeng B, Thurber D, Olson SJ, Alvares CP, Allen ACP, Murphy PD;
PI Critz BS;
XX
DR WPI; 2000-646756/62.
DR N-PSDB; AAC60795.
XX
XX New coding sequence of the human BRCA1 gene, i.e. BRCA1 (om12), useful
PT in gene therapy, especially for preventing or treating breast or
PT ovarian cancer, as well as for diagnosing or monitoring breast or
PT ovarian cancer -
XX
PS Example 4; Column 65-74; 56pp; English.
XX
CC AAC60793 to AAC60795 encode the human BRCA1 (om1-3) proteins given in
CC AAB24217 to AAB24219 respectively. BRCA1 is found on chromosome 17
CC mapping to position 17q21. The BRCA1 (om12) coding sequence is
CC specifically claimed in the present invention. The BRCA1 (om12) coding
CC sequence is useful in gene therapy, especially for preventing or treating
CC breast or ovarian cancer. It is also useful for diagnosing or monitoring
CC breast or ovarian cancer. Furthermore, the BRCA1 (om12) coding sequence
CC is useful for: (a) identifying individuals having BRCA1 gene mutations
CC and having an increased genetic susceptibility to breast or ovarian
CC cancer, or identifying a mutation that increases the genetic
CC susceptibility to breast or ovarian cancer; (b) avoiding
CC misinterpretation of polymorphisms found in the BRCA1 gene; (c)
CC determining the presence of a previously unknown mutation in the BRCA1
CC gene; (d) probing a human sample of the BRCA1 gene by allele to determine
CC the presence of either polymorphic alleles or mutations; and (e)
CC performing diagnosis with a reagent derived from the BRCA1 (om1) cDNA
CC sequence. AAC60796 to AAC60861 represent PCR primers for the BRCA1 gene,
CC which are used in an example from the present invention.
XX
SQ Sequence 1863 AA;
Query Match 100.0%; Score 9649; DB 21; Length 1863;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MDLSALRVEEVQNVINAMQKILECPICIELEIKPEYSTKCDHIFCKFCMLKLLNQKGPQ 60
Db 1 MDLSALRVEEVQNVINAMQKILECPICIELEIKPEYSTKCDHIFCKFCMLKLLNQKGPQ 60
Qy 61 CPLCKNDITKRSLOESTRFSQVLSVEELLKIIICAFQDITGLVANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVLSVEELLKIIICAFQDITGLVANSYNFAKKENNSPEHLKD 120
Qy 121 EVSIIQSMGYRNRAKRLQSEPNPSLOETSLSVOLSNLGTVRTLTQRIOQPKTSVYI 180
Db 121 EVSIIQSMGYRNRAKRLQSEPNPSLOETSLSVOLSNLGTVRTLTQRIOQPKTSVYI 180
Qy 181 ELGSDSSSEDYTNKATYCSVGQDELLOITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSSEDYTNKATYCSVGQDELLOITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ 240
```


QY 241 PSNDLNTTEKRAAERHPKPYQSSVSNLHVPCGNTNTHASSLQHENSNSLLLTQKDRMNVE 300
 DB 241 PSNDLNTTEKRAAERHPKPYQSSVSNLHVPCGNTNTHASSLQHENSNSLLLTQKDRMNVE 300
 QY 301 KAEFCNKSOPGLARSOHNWAGSKETCNDRRTPSTTEKKVLDLADPLCEKEWKKQLPC 360
 DB 301 KAEFCNKSOPGLARSOHNWAGSKETCNDRRTPSTTEKKVLDLADPLCEKEWKKQLPC 360
 QY 361 SENPROTEDEVPIITLNSSIQKVNEWFSRSDDELLGSDSDHGESNAKAVADVLVDLNEVD 420
 DB 361 SENPROTEDEVPIITLNSSIQKVNEWFSRSDDELLGSDSDHGESNAKAVADVLVDLNEVD 420
 QY 421 EYSGSEKIDLLASDPHEALICKSERVHKSVSFNIEDKIFGKYTKKASLPNLSHVTEEN 480
 DB 421 EYSGSEKIDLLASDPHEALICKSERVHKSVSFNIEDKIFGKYTKKASLPNLSHVTEEN 480
 QY 481 LIIGAFVTEPQIIIOERPLTNKLRKRPTSGLHPEDFIKKADLAVOKTPBMIQNGTNOTE 540
 DB 481 LIIGAFVTEPQIIIOERPLTNKLRKRPTSGLHPEDFIKKADLAVOKTPBMIQNGTNOTE 540
 QY 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISSISNNELELNI 600
 DB 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISSISNNELELNI 600
 QY 601 HNSKAPKKNLRKSTRHITHALELVVSRNLSPNCTELQIDSCSSSEEEKKKKNOMPV 660
 DB 601 HNSKAPKKNLRKSTRHITHALELVVSRNLSPNCTELQIDSCSSSEEEKKKKNOMPV 660
 QY 661 RHSRNLQMEGKEPATGAKSKSNPNQOTSKRHSDSDTFPELKLTPNAPGSFTKCSNTSELKE 720
 DB 661 RHSRNLQMEGKEPATGAKSKSNPNQOTSKRHSDSDTFPELKLTPNAPGSFTKCSNTSELKE 720
 QY 721 FVNPSLPREEKEKLETVKYSNNAEDPKDMLSGERVLOTERSVESSSISLVPCTDYGTO 780
 DB 721 FVNPSLPREEKEKLETVKYSNNAEDPKDMLSGERVLOTERSVESSSISLVPCTDYGTO 780
 QY 781 ESTSLLEVSTLGRKAKTEPNKVSQCAAFENPKGLIHGCSKDNNDTEGKYPLGHEVNHHS 840
 DB 781 ESTSLLEVSTLGRKAKTEPNKVSQCAAFENPKGLIHGCSKDNNDTEGKYPLGHEVNHHS 840
 QY 841 RETSIEMESELDQAQYLONTFKVSKQSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900
 DB 841 RETSIEMESELDQAQYLONTFKVSKQSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900
 QY 901 FECEQEENOGKNESNIKPVQTVNITAGFPVVGQKDPVDNAKCSIKGGSRFCLSSQFRG 960
 DB 901 FECEQEENOGKNESNIKPVQTVNITAGFPVVGQKDPVDNAKCSIKGGSRFCLSSQFRG 960
 QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKKNLLEENPEEHSMSPEREMGNENIP 1020
 DB 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKKNLLEENPEEHSMSPEREMGNENIP 1020
 QY 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIQAEGLRNGRPKL 1080
 DB 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIQAEGLRNGRPKL 1080
 QY 1081 NAMRLGVLPQPEVYKQSLPGSNCKHPEIKKQYEEVQTVNTDFSPYLISDNLEQPMGSS 1140
 DB 1081 NAMRLGVLPQPEVYKQSLPGSNCKHPEIKKQYEEVQTVNTDFSPYLISDNLEQPMGSS 1140
 QY 1141 HASQVCSETPDDLLDGEIKEDTSAFENDIKESSAVFSKVQKGLSRSPSPFTHTHLAQ 1200
 DB 1141 HASQVCSETPDDLLDGEIKEDTSAFENDIKESSAVFSKVQKGLSRSPSPFTHTHLAQ 1200
 QY 1201 GYBRGAKKLESSEENLSSDEELPCFQHLFGKVNIPQSOSTRHSTVATECLSKNTEENL 1260
 DB 1201 GYBRGAKKLESSEENLSSDEELPCFQHLFGKVNIPQSOSTRHSTVATECLSKNTEENL 1260
 QY 1261 LSLKNSLNCDSNOVILAKASQEHHLSEETKCSASLFSOCSSELEDLTANTNTQDPFLIGS 1320
 DB 1261 LSLKNSLNCDSNOVILAKASQEHHLSEETKCSASLFSOCSSELEDLTANTNTQDPFLIGS 1320
 QY 1321 SKQMRHQSSEQVGLSKDELVSDEERGTCLENNNOEQSMDSNLGEAASGCESETSVSE 1380

DB 1321 SKQMRHQSSEQVGLSKDELVSDEERGTCLENNNOEQSMDSNLGEAASGCESETSVSE 1380
 QY 1381 DCSGLSSQSDILTTQORDTMOHNLIKLOQEMAELEAVLBOHQSGPSNSYPSIISDSSALE 1440
 DB 1381 DCSGLSSQSDILTTQORDTMOHNLIKLOQEMAELEAVLBOHQSGPSNSYPSIISDSSALE 1440
 QY 1441 DLNRPQSTSEKAVLTQSSESSEYPISONPEGLSADKFEVSADSSTSKNKEPGVERSPSK 1500
 DB 1441 DLNRPQSTSEKAVLTQSSESSEYPISONPEGLSADKFEVSADSSTSKNKEPGVERSPSK 1500
 QY 1501 CPSLDDRWTWYMHSCSGSLQNRNYPQSEELIKVVDVEEQLEESGPHDLTETSYLPRODLEG 1560
 DB 1501 CPSLDDRWTWYMHSCSGSLQNRNYPQSEELIKVVDVEEQLEESGPHDLTETSYLPRODLEG 1560
 QY 1561 TPYLESISILFSDDPSPSEDRAPE SARVGNIPSPSTSAKVPOLKVAESAQSPAAAHFT 1620
 DB 1561 TPYLESISILFSDDPSPSEDRAPE SARVGNIPSPSTSAKVPOLKVAESAQSPAAAHFT 1620
 QY 1621 DTAGYNAMESVSREKPELTASTERNVKNRMSVVSGLTPEEFMLVYKFARKHHITLTNLI 1680
 DB 1621 DTAGYNAMESVSREKPELTASTERNVKNRMSVVSGLTPEEFMLVYKFARKHHITLTNLI 1680
 QY 1681 TEETHVVMKTDAEFVCERTLYFLGIAGGKVVVSYFWVTQSTIKERKMLNEHDFEVRGDV 1740
 DB 1681 TEETHVVMKTDAEFVCERTLYFLGIAGGKVVVSYFWVTQSTIKERKMLNEHDFEVRGDV 1740
 QY 1741 VNGRNHOGKPRARESODRKIFRGLEICCCYGPFTNMPTDQLEWVQLCGASVVKELSSFTL 1800
 DB 1741 VNGRNHOGKPRARESODRKIFRGLEICCCYGPFTNMPTDQLEWVQLCGASVVKELSSFTL 1800
 QY 1801 GTGVHPIVVVQPDAMTEDNGFHAIGOMCEAPVVTREWVLDVSVALYOCQELDTYLIPOIPH 1860
 DB 1801 GTGVHPIVVVQPDAMTEDNGFHAIGOMCEAPVVTREWVLDVSVALYOCQELDTYLIPOIPH 1860
 QY 1861 SHY 1863
 DB 1861 SHY 1863
 RESULT 4
 AAR97128
 ID AAR97128 standard; Protein; 1863 AA.
 XX
 AC AAR97128;
 XX
 DT 19-NOV-1996 (first entry)
 XX
 DE BRCA1, breast and ovarian cancer susceptibility gene product.
 XX
 KW BRCA1; breast cancer; ovary cancer; predisposing gene; diagnosis;
 KW susceptibility gene; prognosis; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO9605308-A1.
 PD
 XX 22-FEB-1996.
 PF 11-AUG-1995; 95WO-US10220.
 XX
 PR 07-JUN-1995; 95US-0488011.
 PR 12-AUG-1994; 94US-0289221.
 PR 02-SEP-1994; 94US-0300266.
 PR 16-SEP-1994; 94US-0308104.
 PR 29-NOV-1994; 94US-0348824.
 PR 24-MAR-1995; 95US-0409305.
 PR 07-JUN-1995; 95US-0483554.
 PR 07-JUN-1995; 95US-0487002.
 XX
 PA (MYRI-) MYRIAD GENETICS INC.
 PA (UTAH) UNIV UTAH RES FOUND.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Futreal PA, Goldgar DE, Harshman KD, Kamb A, Miki Y;
PI Shattuck-Eidens DM, Skolnick MH, Swenson J, Tavtigian SV;
XX Wiseman RW;
XX WPI; 1996-139704/14.
DR N-PSDB; AAT32601.
XX
PT New method for diagnosing a predisposition to breast and ovarian
PT cancer - by detecting a germline alteration in the BRCA1 gene or
PT gene regulatory sequence; for gene therapy and to screen for drugs
XX
PS Claim 12; Page 108-117; 200pp; English.
XX
CC This is the amino acid sequence of the protein encoded by the breast and
CC ovarian cancer susceptibility gene, BRCA1. Four kindred families
CC provided genetic evidence for localisation of BRCA1 to a sufficiently
CC small region for the application of positional cloning strategies. A
CC detailed map of transcripts was developed for the region of 17q21 between
CC D17S1321 and D17S1324. A combination of sequences obt'd. from cDNA
CC clones, hybrid-selected sequences and PCR prods. allowed construction of
CC a composite full-length BRCA1 cDNA (see AAT32612 for genomic sequence).
CC The BRCA1 cDNA revealed an open reading frame encoding a protein of
CC 208 kD. The protein also contains a sequence near the amino terminus
CC which had homology to zinc finger domains, esp. the sequence contains
CC Cys and His residues present as a C3HC4 zinc finger consensus motif.
CC The isolated cDNA is used in methods for either diagnosis of the
CC predisposition to cancer (partic. breast and ovarian cancer), or for the
CC diagnosis or prognosis of cancer, and also in gene-based therapies
CC directed at cancer cells.
XX
SQ Sequence 1863 AA;

Query Match 99.9%; Score 9642; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFMCMLLNQKGP 60
DB 1 MDSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFMCMLLNQKGP 60
QY 61 CPLCKNDITKRSLOSTFRFSOLVBECLKIICAFOLDTGLEYANSYNFAKKENNSPEHLKD 120
DB 61 CPLCKNDITKRSLOSTFRFSOLVBECLKIICAFOLDTGLEYANSYNFAKKENNSPEHLKD 120
QY 121 EYSIIQSMGYRNRRKRLQSEPNPSLOETSLSVOLSNLGTVRTLRKQIQPOKTSVYI 180
DB 121 EYSIIQSMGYRNRRKRLQSEPNPSLOETSLSVOLSNLGTVRTLRKQIQPOKTSVYI 180
QY 181 ELGSDSSEDTYKNATYCSVGDQELIQTPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
DB 181 ELGSDSSEDTYKNATYCSVGDQELIQTPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTEKRAARHPPEKPYCGGSVSNLHVPCGTNTHASSLQHNSSLLLTAKDRNVE 300
DB 241 PSNNDLNTEKRAARHPPEKPYCGGSVSNLHVPCGTNTHASSLQHNSSLLLTAKDRNVE 300
QY 301 KAEFCNKSQPLARSOHNRWAGSKETCNDRTPTSEKKVDNLNADPLCERKENWKQLPC 360
DB 301 KAEFCNKSQPLARSOHNRWAGSKETCNDRTPTSEKKVDNLNADPLCERKENWKQLPC 360
QY 361 SENPRDTEVPWITLNSSIQKVNWFWSRDELGSDSDSHDGESESNKAVADVLVDLNEVD 420
DB 361 SENPRDTEVPWITLNSSIQKVNWFWSRDELGSDSDSHDGESESNKAVADVLVDLNEVD 420
QY 421 EYSGSSEKIDLLASDPHALTCKSERVHSKSVESNIEDKIFGKYTKRKASLPNLSHVTE 480
DB 421 EYSGSSEKIDLLASDPHALTCKSERVHSKSVESNIEDKIFGKYTKRKASLPNLSHVTE 480
QY 481 LIIGAFVTEPQIQRPLTNLKKRRPTSGLHPEDFIKKADLAVOKTPEMINOGTNOTE 540
DB 481 LIIGAFVTEPQIQRPLTNLKKRRPTSGLHPEDFIKKADLAVOKTPEMINOGTNOTE 540

QY 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTESLEKESAFKTKAEPISISSINMELELNI 600
DB 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTESLEKESAFKTKAEPISISSINMELELNI 600
QY 601 HNSKAPKKNRLRRKSSRTHIALELVVSRNLSPNCTELQIDSCSSSEIEIKKKYQNPV 660
DB 601 HNSKAPKKNRLRRKSSRTHIALELVVSRNLSPNCTELQIDSCSSSEIEIKKKYQNPV 660
QY 661 RHSNLIOLMEGKEPATGAKKSNKNEOTSKRHDSTPELKLTNAPGFTKCSNTSELKE 720
DB 661 RHSNLIOLMEGKEPATGAKKSNKNEOTSKRHDSTPELKLTNAPGFTKCSNTSELKE 720
QY 721 FVNPSLPREEKEELETYKVSNNNAEDPKDMLSGERVLTQTERSVESSISLVPGDYGTQ 780
DB 721 FVNPSLPREEKEELETYKVSNNNAEDPKDMLSGERVLTQTERSVESSISLVPGDYGTQ 780
QY 781 ESTISLLEVSTLGKAKTPEKVCVSCAAPENPKGLIHGCSKDNRNDTEGFKYPLGHEVNH 840
DB 781 ESTISLLEVSTLGKAKTPEKVCVSCAAPENPKGLIHGCSKDNRNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMEESLDAQYLQNTFKVSKROSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900
DB 841 RETSIEMEESLDAQYLQNTFKVSKROSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900
QY 901 FECQKEENOGKNESNIKPVQTVNITAGFPVVGQDKPVDNAKCSIKGSRFCLSSQFRG 960
DB 901 FECQKEENOGKNESNIKPVQTVNITAGFPVVGQDKPVDNAKCSIKGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNLEENFEHSHMSPEREMGENIP 1020
DB 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNLEENFEHSHMSPEREMGENIP 1020
QY 1021 STVSTISRNINRENVFKEASSNINEVGSSSTNEIGSSDENTQAEILGNRGPKL 1080
DB 1021 STVSTISRNINRENVFKEASSNINEVGSSSTNEIGSSDENTQAEILGNRGPKL 1080
QY 1081 NAMLRGLVLOPEVYKQSLPGSNCKHPETKQYEEVQTVNTDFSPYILSDNLEQPMSS 1140
DB 1081 NAMLRGLVLOPEVYKQSLPGSNCKHPETKQYEEVQTVNTDFSPYILSDNLEQPMSS 1140
QY 1141 HASQVCSQETPDDLLDGEIKEDTSAENDIKESSAVFSKSVQKGLSRSPFTHTHAQ 1200
DB 1141 HASQVCSQETPDDLLDGEIKEDTSAENDIKESSAVFSKSVQKGLSRSPFTHTHAQ 1200
QY 1201 GYRGAKKLESSEENLSSEDEELPCFOHLLFGKVNNPISOSTRUSTVATECLSKNTEENL 1260
DB 1201 GYRGAKKLESSEENLSSEDEELPCFOHLLFGKVNNPISOSTRUSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNQVILAKASQEHLSSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
DB 1261 LSLKNSLNDCSNQVILAKASQEHLSSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHOSQSGVGLSDKELVSDDEERGTLGLENNQBEQSMDSNLGEAASCESTSVSE 1380
DB 1321 SKQMRHOSQSGVGLSDKELVSDDEERGTLGLENNQBEQSMDSNLGEAASCESTSVSE 1380
QY 1381 DCSGLSSQSDILTTQQRDTHQHNLIKLQOEAALEAVLEQHGQSPNSYPSITIDSSSALE 1440
DB 1381 DCSGLSSQSDILTTQQRDTHQHNLIKLQOEAALEAVLEQHGQSPNSYPSITIDSSSALE 1440
QY 1441 DLNRPEQSTSEKAVLTSQKSSEYPIQNPGLSADKFEVSADSSSTSKNKEPVERSSPSK 1500
DB 1441 DLNRPEQSTSEKAVLTSQKSSEYPIQNPGLSADKFEVSADSSSTSKNKEPVERSSPSK 1500
QY 1501 CPSLDDRWYMHSCGSLQNRNYPQBELIKVVDVEEQLEESGPHDLTETSYLPQDLEG 1560
DB 1501 CPSLDDRWYMHSCGSLQNRNYPQBELIKVVDVEEQLEESGPHDLTETSYLPQDLEG 1560
QY 1561 TPYLESGISLFSDDPSPEDRAPESARVGNIPSSTSALKVPOLKVAESQAQSPAAHTT 1620
DB 1561 TPYLESGISLFSDDPSPEDRAPESARVGNIPSSTSALKVPOLKVAESQAQSPAAHTT 1620
QY 1621 DTAGYNAMESVSRKPELPTASTERVNRKRMVMVSGLTPEEFMLVYKPFARKHHTLTNLI 1680

```
|||||
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSMVVSGLTPEEFMLVYKFARKHHTITNLII 1680
QY 1681 TETTHVVMKTDAEFCERTLKVELGIAGGKVVSVFWVTQSTKERKMLNEHDFEVRGDV 1740
Db 1681 TETTHVVMKTDAEFCERTLKVELGIAGGKVVSVFWVTQSTKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQGPKRARESODRKIFRGLGICCYGPTNMPDQLEWMVQLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQGPKRARESODRKIFRGLGICCYGPTNMPDQLEWMVQLCGASVVKELSSFTL 1800
QY 1801 GTGVHPVIVVQPDADWETDNGFHAIGMCBAPVVTREWVLDVALYQCQELDTYLIPOIPH 1860
Db 1801 GTGVHPVIVVQPDADWETDNGFHAIGMCBAPVVTREWVLDVALYQCQELDTYLIPOIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863
|||
RESULT 5
ID AAR81481 standard; Protein; 1863 AA.
AC AAR81481;
XX
DT 30-SEP-1996 (first entry)
XX
DE BRCA1.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX
OS Homo sapiens.
XX
PN WO9605306-A2.
XX
PD 22-FEB-1996.
XX
PF 11-AUG-1995; 95WO-US10202.
XX
PR 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0480784.
XX
PA (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.
XX
PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
XX
PI WPI; 1996-139702/14.
XX
DR N-PSDB; AAT17438.
XX
PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
XX
PS Claim 1; Page 119-128; 218pp; English.
XX
CC This sequence represents the protein encoded by the human breast and
CC ovarian cancer predisposing gene (BRCA1). Mutations of this sequence
CC (see AAR81483-R81497 and AAR81499-R81546) can be used as immunogens for
CC antibody production. The mutant BRCA1 genes have at least 1 mutation or
CC polymorphism in comparison to the cDNA encoding this sequence. By
CC detecting a germline alteration in this gene, a predisposition for
CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
CC isolated from a tissue sample from a subject has a probe, corresponding
```

```
.CC to a fragment of the cDNA encoding this sequence (or an allele-specific
CC probe for a mutation of it sequence), added to it. The conditions allow
CC for hybridisation of the probe to the mRNA, and any hybridisation which
CC occurs is detected. Alternatively the BRCA1 gene in the tissue sample
CC is isolated, and a shift in electrophoretic mobility of single stranded
CC DNA from the sample on a non-denaturing polyacrylamide gel indicates a
CC mutation. These methods of detection can also diagnose a lesion
CC neoplasia associated with the BRCA1 locus. The methods may be used in
CC gene therapy, protein replacement therapy and protein mimetics, and may
CC be used to screen for drugs in cancer therapy.
XX
SQ Sequence 1863 AA;
Query Match 99.9%; Score 9642; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEEVQNVINAMQKILECPICLEIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ 60
Db 1 MDLSALRVEEVQNVINAMQKILECPICLEIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ 60
QY 61 CPLCKNDITKRSLOESTRESQVLELLKIIICAFQDITGLYANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRESQVLELLKIIICAFQDITGLYANSYNFAKKENNSPEHLKD 120
QY 121 EYSIIQSMGYRNRAKRLLOSEPENPSLOETSLSVQLSNLGTVRTLTORKIORIQKTSVYI 180
Db 121 EYSIIQSMGYRNRAKRLLOSEPENPSLOETSLSVQLSNLGTVRTLTORKIORIQKTSVYI 180
QY 181 ELGSDSSEDTVNKATYCVGDQELLOITPQGRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDTVNKATYCVGDQELLOITPQGRDEISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVPCGTNTHASSLOHENSLLLTDKRMNVE 300
Db 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVPCGTNTHASSLOHENSLLLTDKRMNVE 300
QY 301 KAEFCNKSKQPGIARSOHNRWAGSKETCNDRRTPSTPEKKVVDLNADPLCKERKNWKQLPC 360
Db 301 KAEFCNKSKQPGIARSOHNRWAGSKETCNDRRTPSTPEKKVVDLNADPLCKERKNWKQLPC 360
QY 361 SENPRDTEVPWITLNSSIQKYNWFPSRDELGLSDSDSHDGESESNKAVADVLDVLENYD 420
Db 361 SENPRDTEVPWITLNSSIQKYNWFPSRDELGLSDSDSHDGESESNKAVADVLDVLENYD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYKKASLNLNLSHTEN 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYKKASLNLNLSHTEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKKADLAVOKTPEMINOQTNQTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKKADLAVOKTPEMINOQTNQTE 540
QY 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSISNMELELNI 600
Db 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSISNMELELNI 600
QY 601 HNSKAPKKNLRLRKSSTRHIALELVVSRNLSPPPNCTELOIDSCSSSEIKKKKYNQMPV 560
Db 601 HNSKAPKKNLRLRKSSTRHIALELVVSRNLSPPPNCTELOIDSCSSSEIKKKKYNQMPV 560
QY 661 RUSRNQLMEGKEPATGAKKSNKPNEOTSKRHDSDFPELKLNTNAPGFTKCSNTSELKE 720
Db 661 RUSRNQLMEGKEPATGAKKSNKPNEOTSKRHDSDFPELKLNTNAPGFTKCSNTSELKE 720
QY 721 FYNPSLPREEKEKLETVKVSNNAEDPKDMLSGERVLQTERSVESSSISLVPDGYGTQ 780
Db 721 FYNPSLPREEKEKLETVKVSNNAEDPKDMLSGERVLQTERSVESSSISLVPDGYGTQ 780
QY 781 ESISLLEVSTLGAKEPKNKCVSQAAPENPKGLIHGCSKDNRNDETEGPKYPLGHEVNHS 840
Db 781 ESISLLEVSTLGAKEPKNKCVSQAAPENPKGLIHGCSKDNRNDETEGPKYPLGHEVNHS 840
```

QY 841 RETSIEMESELDAQYLQNTFKVSKROSFALFSPNGNAEECAATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSKROSFALFSPNGNAEECAATFSAHSGSLKKQSPKVT 900
QY 901 FCEQKEENOGKNESNIKPQVTNITAGFPVVGOKDKPVDNAKSIKGGSRFLSSQFRG 960
Db 901 FCEQKEENOGKNESNIKPQVTNITAGFPVVGOKDKPVDNAKSIKGGSRFLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPRIPLFPPIKSFVTKCKKNLLEENFEHSMSPEREMGNENIP 1020
Db 961 NETGLITPNKHGLLQNPRIPLFPPIKSFVTKCKKNLLEENFEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNNIRENVFKKASSNNINEVGSSSTNEIGSSDENEIQABELGRNRPGL 1080
Db 1021 STVSTISRNNIRENVFKKASSNNINEVGSSSTNEIGSSDENEIQABELGRNRPGL 1080
QY 1081 NAMLRGLVQLPEVYKQSLPGSNCKHPEIKKOEYEEVQTVNTDFSPYLIISDNLEQPMGSS 1140
Db 1081 NAMLRGLVQLPEVYKQSLPGSNCKHPEIKKOEYEEVQTVNTDFSPYLIISDNLEQPMGSS 1140
QY 1141 HASOVCSSTPDLLDDGGEIKEDTSPAENDIKESSAVFSKSVOKGELSRSPFTHHLAQ 1200
Db 1141 HASOVCSSTPDLLDDGGEIKEDTSPAENDIKESSAVFSKSVOKGELSRSPFTHHLAQ 1200
QY 1201 GYRRGAKKLESSENLSEDEELPCFQHLFLGKYNNIPSPQSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSENLSEDEELPCFQHLFLGKYNNIPSPQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLDCSNQVTLAKASQEHLSSEETKCSASFSSQCSSELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLDCSNQVTLAKASQEHLSSEETKCSASFSSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHOSOGVGLSDKELVSDDEERGTLGEENNOEQSMDSNLGEAASCESETSVSE 1380
Db 1321 SKQMRHOSOGVGLSDKELVSDDEERGTLGEENNOEQSMDSNLGEAASCESETSVSE 1380
QY 1381 DCSGLSSQSDILTTQQRDTMHOHNIKIQEAMAEAVLEQHGQSPNSYPSIISDSSALE 1440
Db 1381 DCSGLSSQSDILTTQQRDTMHOHNIKIQEAMAEAVLEQHGQSPNSYPSIISDSSALE 1440
QY 1441 DURNPEQSTSEKAVLTSQKSEYPIQONPEGLSADKFEVSADSSSTKKNKPGVERSSPSK 1500
Db 1441 DURNPEQSTSEKAVLTSQKSEYPIQONPEGLSADKFEVSADSSSTKKNKPGVERSSPSK 1500
QY 1501 CPSLDDRYMHSCSGSLQNRNYPQOEELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDDRYMHSCSGSLQNRNYPQOEELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESGLISLSDPESDPSDRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT 1620
Db 1561 TPYLESGLISLSDPESDPSDRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPTPEFMLVYKFAKKHITLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPTPEFMLVYKFAKKHITLTNLI 1680
QY 1681 TEETHVVMKTDAPVCERTLKYFLGTAGGKVVVSYFWVTQSTIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDAPVCERTLKYFLGTAGGKVVVSYFWVTQSTIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQPKKRAESQDRKIFRGLIEICCYGPFNTNMPTDQLEWMVQLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQPKKRAESQDRKIFRGLIEICCYGPFNTNMPTDQLEWMVQLCGASVVKELSSFTL 1800
QY 1801 GTGVHPIVVVQPDWEDNGFHAIGQWCEAPVVTREWVLDVSVALYQCQELDTYLIPOIPH 1860
Db 1801 GTGVHPIVVVQPDWEDNGFHAIGQWCEAPVVTREWVLDVSVALYQCQELDTYLIPOIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 6
AAR91208
ID AAR91208 standard; Protein; 1863 AA.
XX AAR91208;
AC AAR91208;
XX 05-JUN-1996 (first entry)
DT
XX BRCA1, breast and ovarian cancer predisposing gene product.
DE
XX BRCA1; breast cancer; ovary cancer; predisposing gene;
KW susceptibility gene; protein replacement therapy; diagnosis;
KW prognosis.
XX Homo sapiens.
OS
XX W09605307-A2.
PN
XX 22-FEB-1996.
PD
XX 11-AUG-1995; 95WO-US10203.
PF
XX 07-JUN-1995; 95US-0488011.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0483554.
PR 07-JUN-1995; 95US-0487002.
XX (MYRI-) MYRIAD GENETICS INC.
PA (UTAH) UNIV UTAH RES FOUND.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX Futreal AP, Goldgar DE, Harshman KD, Kamb A, Miki Y;
PI Shattuck-Eidens DM, Skolnick MH, Swenson J, Tavtigian SV;
PI Wiseman RW;
XX WPI; 1996-139703/14.
DR N-PSDB; AAT18310.
XX New isolated human cancer predisposing gene, BRCA1 - used to develop
PT prods. for diagnosis, prognosis and therapy of cancers, partic.
PT breast and ovarian cancers
XX
PS Claim 1; Page 117-124; 190pp; English.
XX
CC The BRCA1 polypeptide (AAR91208) is the product of the human BRCA1
CC gene (AAT18310). It can be obtd. by insertion of BRCA1 DNA into a
CC vector and expression in host cells; products of mutated BRCA1
CC genes, associated in humans with a predisposition to breast and
CC ovarian cancer, can also be produced. BRCA1 polypeptides are used
CC as immunogens for the prodn. of antibodies, or to screen therapeutic
CC drugs. They can also be used in protein replacement therapy.
XX
SQ Sequence 1863 AA;
Query Match 99.9%; Score 9642; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEEVQNVINAMOKILECPICILELKEPVTCKDHIFCKFCMLKLLNOKKGPSQ 60
Db 1 MDLSALRVEEVQNVINAMOKILECPICILELKEPVTCKDHIFCKFCMLKLLNOKKGPSQ 60
QY 61 CPLCKNDITKRSLOESTRFSOLVVEELIKIICAFOLDTGLEYANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSOLVVEELIKIICAFOLDTGLEYANSYNFAKKENNSPEHLKD 120
QY 121 EVSIIOGMYNRNRAKRLQSEPNPSLQETSLSVOLSNLGTVRTLRTKQRTQPKTSVYI 180
Db 121 EVSIIOGMYNRNRAKRLQSEPNPSLQETSLSVOLSNLGTVRTLRTKQRTQPKTSVYI 180

QY	181	ELGSDSSEDTVVKATYCSVGQDQLLIQTQGTREDEISLDSAKKAACEFSETDVTNTEHHQ	240
Db	181	ELGSDSSEDTVVKATYCSVGQDQLLIQTQGTREDEISLDSAKKAACEFSETDVTNTEHHQ	240
QY	241	PSNNDLNTTEKRAAERHPKPYQGSYSVNLHVPCGCTNTIASSLQHENSLLLTLDKRMNVE	300
Db	241	PSNNDLNTTEKRAAERHPKPYQGSYSVNLHVPCGCTNTIASSLQHENSLLLTLDKRMNVE	300
QY	301	KAEFCNKSQKQGLARSHNRWAGSKETCNDRRTPSTPEKKVVDLNADPLCERKEWNKOKLPC	360
Db	301	KAEFCNKSQKQGLARSHNRWAGSKETCNDRRTPSTPEKKVVDLNADPLCERKEWNKOKLPC	360
QY	361	SENPRDTEVPWITLNSSIQKYNWFSRSDDELIGSDSDHDSGESENAKVADVLVDLNEVD	420
Db	361	SENPRDTEVPWITLNSSIQKYNWFSRSDDELIGSDSDHDSGESENAKVADVLVDLNEVD	420
QY	421	EYSGSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKYRKKASLPLNSHVTE	480
Db	421	EYSGSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKYRKKASLPLNSHVTE	480
QY	481	LIIGAEVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKKADLAVOKTPEMINOQNTOTE	540
Db	481	LIIGAEVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKKADLAVOKTPEMINOQNTOTE	540
QY	541	QNGQVNNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI	600
Db	541	QNGQVNNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI	600
QY	601	HNSKAPKNRLRKSTRIHHALELVVSRNLSPPNCTELQIDSCSSSEIEIKKKYNQMPV	660
Db	601	HNSKAPKNRLRKSTRIHHALELVVSRNLSPPNCTELQIDSCSSSEIEIKKKYNQMPV	660
QY	661	RISRNQLMEGKEPATGAKSKNPNQETSKRHDSDFPELKLNTAPGCTFKCSNTSELKE	720
Db	661	RISRNQLMEGKEPATGAKSKNPNQETSKRHDSDFPELKLNTAPGCTFKCSNTSELKE	720
QY	721	FVNPSLPREEKEEKLTVKYSVNAEDPKDMLLSGERVLQTERSVESSISLVPGTDYGTQ	780
Db	721	FVNPSLPREEKEEKLTVKYSVNAEDPKDMLLSGERVLQTERSVESSISLVPGTDYGTQ	780
QY	781	ESISLLEVSTLGAKEPKNKYSCQAAPENPKGLIHGCKDNRNDETEGFKYPLGHEVNIH	840
Db	781	ESISLLEVSTLGAKEPKNKYSCQAAPENPKGLIHGCKDNRNDETEGFKYPLGHEVNIH	840
QY	841	RETSTEMESELDAQYLQNTFKVSKROSALFSPNPGNAEECATFSAHSGSLKKQSPKVT	900
Db	841	RETSTEMESELDAQYLQNTFKVSKROSALFSPNPGNAEECATFSAHSGSLKKQSPKVT	900
QY	901	FCEQKEENOGKNESNIKPVQTVNITAGPPVVGQDKPVDNAKCSIKGGSRFCLSSQFRG	960
Db	901	FCEQKEENOGKNESNIKPVQTVNITAGPPVVGQDKPVDNAKCSIKGGSRFCLSSQFRG	960
QY	961	NETGLITPNKHGLLQNPYRIPPLPTKSFVKYCKKNLLBENFEHSMSPEREMGNENIP	1020
Db	961	NETGLITPNKHGLLQNPYRIPPLPTKSFVKYCKKNLLBENFEHSMSPEREMGNENIP	1020
QY	1021	STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNRPKL	1080
Db	1021	STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNRPKL	1080
QY	1081	NAMLRLGVLOPEYVKQSLPGSNCKHPEIKKQYEEVYQTVNTDFSPYLLSDNLEQPMGSS	1140
Db	1081	NAMLRLGVLOPEYVKQSLPGSNCKHPEIKKQYEEVYQTVNTDFSPYLLSDNLEQPMGSS	1140
QY	1141	HASQVCSETPDLLDDGETIKEDTSAENDIKESSAVFSKVQKGLSRSPSFTHHLAQ	1200
Db	1141	HASQVCSETPDLLDDGETIKEDTSAENDIKESSAVFSKVQKGLSRSPSFTHHLAQ	1200
QY	1201	GYRRGAKKLLESSEENLSSDEELPCFQHLFGKVNNIQPSQSTRHSTVATECLSKNTEENL	1260
Db	1201	GYRRGAKKLLESSEENLSSDEELPCFQHLFGKVNNIQPSQSTRHSTVATECLSKNTEENL	1260

QY	1261	LSLKNSLNCDSNQVILAKAQEHLSSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS	1320
Db	1261	LSLKNSLNCDSNQVILAKAQEHLSSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS	1320
QY	1321	SKOMRHQSESQGVGLSDKELVSDDEERGTLGLENNOEBSQMSNLGEAASGCESETSVSE	1380
Db	1321	SKOMRHQSESQGVGLSDKELVSDDEERGTLGLENNOEBSQMSNLGEAASGCESETSVSE	1380
QY	1381	DCSGLSSQSDILTTQQRDTMOHNLIKLOQEMAELEAVLQHQSGOPSNSVPSIIISDSALLE	1440
Db	1381	DCSGLSSQSDILTTQQRDTMOHNLIKLOQEMAELEAVLQHQSGOPSNSVPSIIISDSALLE	1440
QY	1441	DLRNPQSTSEKAVLTSQKSSEYPIISQNPESGLSADKFEVSADSTSKNKEPGVERSSPSK	1500
Db	1441	DLRNPQSTSEKAVLTSQKSSEYPIISQNPESGLSADKFEVSADSTSKNKEPGVERSSPSK	1500
QY	1501	CPSLDRRWYHSCSGSLQNRNYPQSEELIKVVDVVEQQLLEESGPHDLTETSYLPRODLEG	1560
Db	1501	CPSLDRRWYHSCSGSLQNRNYPQSEELIKVVDVVEQQLLEESGPHDLTETSYLPRODLEG	1560
QY	1561	TPYLESIGISLFSDDPESDSEDRAPEARVGNIPSPSTSALKVPOLKVAESAQSPAAAHIT	1620
Db	1561	TPYLESIGISLFSDDPESDSEDRAPEARVGNIPSPSTSALKVPOLKVAESAQSPAAAHIT	1620
QY	1621	DRAGYNAMESVSREKPELTASTERYNKRMSVYVSGLTPEEFMLVYKFARKHHITLTNLI	1680
Db	1621	DRAGYNAMESVSREKPELTASTERYNKRMSVYVSGLTPEEFMLVYKFARKHHITLTNLI	1680
QY	1681	TEETTHVVKMTDAEFVCERTLKYFLGIAGGKVVVSVFWTQTSIKERKMLNEHDFEVRGDV	1740
Db	1681	TEETTHVVKMTDAEFVCERTLKYFLGIAGGKVVVSVFWTQTSIKERKMLNEHDFEVRGDV	1740
QY	1741	VNGRNHOGKPRARESODRKIFRGLEICCYGPTNMPDTQLEWVQLCGASVVKELSSFTL	1800
Db	1741	VNGRNHOGKPRARESODRKIFRGLEICCYGPTNMPDTQLEWVQLCGASVVKELSSFTL	1800
QY	1801	GTGVHPVIVVQPDAMWEDNGFHAIGQCEAPVVTREWLDSVALYQCQLDVTYLIPQIPH	1860
Db	1801	GTGVHPVIVVQPDAMWEDNGFHAIGQCEAPVVTREWLDSVALYQCQLDVTYLIPQIPH	1860
QY	1861	SHY 1863	
Db	1861	SHY 1863	
RESULT 7			
AAY32033			
ID	AAY32033	standard; protein; 1863 AA.	
XX			
AC	AAY32033;		
XX			
DT	05-JAN-2000	(first entry)	
XX			
DE	Human BRCA1 protein.		
XX			
KW	BRCA1; p53 protein; p21 gene; human; tumour suppressor;		
KW	transcriptional activator; breast cancer; cell proliferation;		
KW	apoptosis; diagnosis; anticancer; antitumour; drug screening.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	Binding-site	224..500	
FT		/note= "binding site of p53 protein"	
FT	Region	499..510	
FT		/note= "NLS region"	
XX			
PN	WO9950280-A1.		
XX			
PD	07-OCT-1999.		
XX			
PF	31-MAR-1999;	99WO-US07150.	
XX			

PR 31-MAR-1998; 98US-0080146.
XX (UYPE-) UNIV PENNSYLVANIA.
XX El-Deiry WS, Weber BL;
XX WPI; 1999-601319/51.
XX Nucleic acid involved in BRCA-1-mediated control of transcriptional
PT regulation of tumour suppressor genes and related peptides, used to
PT screen for modulators for use as anticancer agents -
XX
PS Claim 9; Page 76-80; 94pp; English.
XX This sequence represents the human BRCA1 protein. BRCA-1
CC inhibits S-phase cell-cycle progression by transactivating
CC expression of p21 in a p53-independent manner. It also
CC interacts with p53 (see AA32034) in vitro and in vivo, and
CC increases p53-dependent transcription from the p21 (see AA220131)
CC and bax promoters. These newly discovered pathways of BRCA1 action
CC provide novel targets to which pharmaceutical agents capable of
CC affecting cell proliferation and apoptosis can be isolated, and
CC can then be used in the treatment and control of cellular
CC proliferation disorders, such as breast cancer. Assays and
CC compositions for identifying compounds that enhance or repress
CC cellular proliferation via these BRCA1-mediated pathways are
CC disclosed.
XX
SQ Sequence 1863 AA;
Query Match 99.9%; Score 9642; DB 20; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEEONVINAMQKILECPICILELKEPVTSCDHFCKFCKMLLNKKGPSQ 60
DB 1 MDLSALRVEEONVINAMQKILECPICILELKEPVTSCDHFCKFCKMLLNKKGPSQ 60
QY 61 CPLCNDITKRSLOESTRFSOLVPELLKICAFOLDTGLEVANSYNFAKKENNSPEHLKD 120
DB 61 CPLCNDITKRSLOESTRFSOLVPELLKICAFOLDTGLEVANSYNFAKKENNSPEHLKD 120
QY 121 EYSIIQSGMYRNRKRLLOSEPNSLQETSLSVQLSNLGTVRTLRTKRIQPKQTSVYI 180
DB 121 EYSIIQSGMYRNRKRLLOSEPNSLQETSLSVQLSNLGTVRTLRTKRIQPKQTSVYI 180
QY 181 ELGSDSSEDTVNKATYCSVGQDQLLOITPQGTREISLDSAKKAACEFSETDVTNTEHQ 240
DB 181 ELGSDSSEDTVNKATYCSVGQDQLLOITPQGTREISLDSAKKAACEFSETDVTNTEHQ 240
QY 241 PSNNDLNTTEKRAERHPEKYQGSVSNLHVPCGTNTHASSLQHNSSLLLTKDRMVE 300
DB 241 PSNNDLNTTEKRAERHPEKYQGSVSNLHVPCGTNTHASSLQHNSSLLLTKDRMVE 300
QY 301 KAEFCNKSQPLARSOHRNAGSKETCNDRRTPSTKVKVDLADPLCERKWNKQKLPC 360
DB 301 KAEFCNKSQPLARSOHRNAGSKETCNDRRTPSTKVKVDLADPLCERKWNKQKLPC 360
QY 361 SENPRDTEVPWITLNSIQKWNWFSSRDELGSDSDSHGSESNAKVADVLVNEVD 420
DB 361 SENPRDTEVPWITLNSIQKWNWFSSRDELGSDSDSHGSESNAKVADVLVNEVD 420
QY 421 EYSGSSEKIDLLADPHALICKSERVHKSVESNIEDKIFGKYRKASLPNLSHVTEN 480
DB 421 EYSGSSEKIDLLADPHALICKSERVHKSVESNIEDKIFGKYRKASLPNLSHVTEN 480
QY 481 LIIGAFVTEPQIIQERPLTNLKRKRRTSGLHPEDFIKKADLAVQKTPEMINGTNOTE 540
DB 481 LIIGAFVTEPQIIQERPLTNLKRKRRTSGLHPEDFIKKADLAVQKTPEMINGTNOTE 540
QY 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELENI 600
DB 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELENI 600

QY 601 HNSKAPKKNRLRRKSSSTRHIAHALLVVSRNLSPNCTELOIDSCSSSEBEIKKKKYNOMPV 660
DB 601 HNSKAPKKNRLRRKSSSTRHIAHALLVVSRNLSPNCTELOIDSCSSSEBEIKKKKYNOMPV 660
QY 661 RHSRNLQLMGKKEPATGAKKSNKPNQTSKRHSDTTFPELKLITNAPGSFTKCSNTSELKE 720
DB 661 RHSRNLQLMGKKEPATGAKKSNKPNQTSKRHSDTTFPELKLITNAPGSFTKCSNTSELKE 720
QY 721 FVNPSLPREKEEKLETVKVSNNADPKDMLSGSERVLQTSRVSSESSISLVPGTDYGTQ 780
DB 721 FVNPSLPREKEEKLETVKVSNNADPKDMLSGSERVLQTSRVSSESSISLVPGTDYGTQ 780
QY 781 ESIISLLEVTSLGKAKTEPNKCVSQAAPENPKGLTHGCSKONRNDTEGFKYPLGHEVNHS 840
DB 781 ESIISLLEVTSLGKAKTEPNKCVSQAAPENPKGLTHGCSKONRNDTEGFKYPLGHEVNHS 840
QY 841 RETSIEMEESLDAQYLQNTFKVSKROSFAFSPNGNAEEECATFSAHSGSLKKQSPKVT 900
DB 841 RETSIEMEESLDAQYLQNTFKVSKROSFAFSPNGNAEEECATFSAHSGSLKKQSPKVT 900
QY 901 FECQKEENQKKNESNKPVTQVNTITAGFPVVGOKDKPVDNAKCSIKGSRFCLSSQPRG 960
DB 901 FECQKEENQKKNESNKPVTQVNTITAGFPVVGOKDKPVDNAKCSIKGSRFCLSSQPRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKNLLLEENFEHSMSPEREMGENIP 1020
DB 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKNLLLEENFEHSMSPEREMGENIP 1020
QY 1021 STVSTISRNINRENVFKEASSNINEVGSSNINEGSSINEIGSSDENIQAEILGRNRGPKL 1080
DB 1021 STVSTISRNINRENVFKEASSNINEVGSSNINEGSSINEIGSSDENIQAEILGRNRGPKL 1080
QY 1081 NAMURLGVLOPEVYKQSLPGSNCKHPEIKQOEYEVQTVNTDFSPYLISNLEOPMGSS 1140
DB 1081 NAMURLGVLOPEVYKQSLPGSNCKHPEIKQOEYEVQTVNTDFSPYLISNLEOPMGSS 1140
QY 1141 HASQVCSCTPDDLDDGEIKEEDTSAENDIKESSAVFSKVQKGBLSRSPFTTHLAAQ 1200
DB 1141 HASQVCSCTPDDLDDGEIKEEDTSAENDIKESSAVFSKVQKGBLSRSPFTTHLAAQ 1200
QY 1201 GYRGAKKLESSEENLSEDEELPCFQHLFLGKVNNIPSQSTRHSTVATECLSKNTEENL 1260
DB 1201 GYRGAKKLESSEENLSEDEELPCFQHLFLGKVNNIPSQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNQVILAKASOEHLHSEETKCSASFSSQCSSELEDLTANTQTDPFLTGS 1320
DB 1261 LSLKNSLNDCSNQVILAKASOEHLHSEETKCSASFSSQCSSELEDLTANTQTDPFLTGS 1320
QY 1321 SKOMRHQSESQGVGLSDKELVSDDEERGTLGEENNOEBSMDNLGEAASCESTSVSE 1380
DB 1321 SKOMRHQSESQGVGLSDKELVSDDEERGTLGEENNOEBSMDNLGEAASCESTSVSE 1380
QY 1381 DCSGLSSQSDILTTQORDTMOHNLIKLQOEAMAEALVLEQHGQSPSNSYPSIISDSSALE 1440
DB 1381 DCSGLSSQSDILTTQORDTMOHNLIKLQOEAMAEALVLEQHGQSPSNSYPSIISDSSALE 1440
QY 1441 DLNRPEOSTSEKAVLTQKSSEYPISONPEGLSADKFEVSADSTSKNKEPGEVERSSPSK 1500
DB 1441 DLNRPEOSTSEKAVLTQKSSEYPISONPEGLSADKFEVSADSTSKNKEPGEVERSSPSK 1500
QY 1501 CPSLDDRWYHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPQDLEG 1560
DB 1501 CPSLDDRWYHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPQDLEG 1560
QY 1561 TPYLESGISLFSDDPESDPESTRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT 1620
DB 1561 TPYLESGISLFSDDPESDPESTRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVVSGLTPEEFMLVYKFAKHHTLTNLI 1680
DB 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVVSGLTPEEFMLVYKFAKHHTLTNLI 1680

QY 1681 TEETHVVMKTDABFVCERTLKTYFLGIAGGKVVVSYFWVTQSTIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDABFVCERTLKTYFLGIAGGKVVVSYFWVTQSTIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHOGPKRARESQRDKIFRGLIEICYGPFTNMPDQLEWVQLGASVVKELSSFTL 1800
Db 1741 VNGRNHOGPKRARESQRDKIFRGLIEICYGPFTNMPDQLEWVQLGASVVKELSSFTL 1800
QY 1801 GTGVHPVVVQPDWEDNGFHAIGQMCEAPVVTREWVLDSVALYQCQELDTYLIPQIPH 1860
Db 1801 GTGVHPVVVQPDWEDNGFHAIGQMCEAPVVTREWVLDSVALYQCQELDTYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863
RESULT 8
ABG01696
ID ABG01696 standard; Protein; 2353 AA.
XX AC ABG01696;
DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #1687.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS65883.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 32055; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 2353 AA;
Query Match 99.9%; Score 9642; DB 22; Length 2353;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEEVQNVINAMQKILECPICLELIKPEVSKCDHIKFCMQLLNQKKGPSQ 60
Db 491 MDLSALRVEEVQNVINAMQKILECPICLELIKPEVSKCDHIKFCMQLLNQKKGPSQ 550
QY 61 CPLCKNDITKRSLOESTRFSQVLELLKIICAFOLDGTGLEVANSYNFAKKENNSPEHLKD 120
Db 551 CPLCKNDITKRSLOESTRFSQVLELLKIICAFOLDGTGLEVANSYNFAKKENNSPEHLKD 610
QY 121 EVSTIQSMGYRNRAKRLQLQSEPNPSLQETSLSVQLSNLGTVRTLRKQIQOKTSVYI 180
Db 611 EVSTIQSMGYRNRAKRLQLQSEPNPSLQETSLSVQLSNLGTVRTLRKQIQOKTSVYI 670
QY 181 ELGSDSSEDTVNKATYCSVGDOELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 671 ELGSDSSEDTVNKATYCSVGDOELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 730
QY 241 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGTNTHASSLQHENSLLLTKDRMNVE 300
Db 731 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGTNTHASSLQHENSLLLTKDRMNVE 790
QY 301 KAEFCNKSQKPLARSQHNWAGSKETCNDRRTPSTPEKKVLDNADPLCERKENWKQLPC 360
Db 791 KAEFCNKSQKPLARSQHNWAGSKETCNDRRTPSTPEKKVLDNADPLCERKENWKQLPC 850
QY 361 SENPRDTEVPWITLANSIOKVNEWFSRDELGSDSDHGESESNAAKVADVDLVNEVD 420
Db 851 SENPRDTEVPWITLANSIOKVNEWFSRDELGSDSDHGESESNAAKVADVDLVNEVD 910
QY 421 EYSGSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKYRKKASLPNLSHVTEN 480
Db 911 EYSGSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKYRKKASLPNLSHVTEN 970
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPEMINOGTNOTE 540
Db 971 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPEMINOGTNOTE 1030
QY 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTAEPISSSISNMELELNI 600
Db 1031 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTAEPISSSISNMELELNI 1090
QY 601 HNSKAPKKNLRKRSSTRHIALELVVSRNLSPNPNCTELOIDSCSSSEEEKKKKYNOMPV 660
Db 1091 HNSKAPKKNLRKRSSTRHIALELVVSRNLSPNPNCTELOIDSCSSSEEEKKKKYNOMPV 1150
QY 661 RHRSNQLMEGKEPATGAKKSNKPNQETSKRHSDTPELKLTNAPGFTKCSNTSELKE 720
Db 1151 RHRSNQLMEGKEPATGAKKSNKPNQETSKRHSDTPELKLTNAPGFTKCSNTSELKE 1210
QY 721 FVNPSPREKEKEKLETVKVNNAEDPKDMLSGERVLOQTORSVESSSISLVPGTDYGTQ 780
Db 1211 FVNPSPREKEKEKLETVKVNNAEDPKDMLSGERVLOQTORSVESSSISLVPGTDYGTQ 1270
QY 781 ESTSLLVSVTLGAKTPEPKVSCAAAFENPKGLIHGCSKDNDRNDTGFKYPLGHEVNH 840
Db 1271 ESTSLLVSVTLGAKTPEPKVSCAAAFENPKGLIHGCSKDNDRNDTGFKYPLGHEVNH 1330
QY 841 RETSIEMESELDAQYLQNTFKVSKROSFALFNPNGNAEBECATFSAHSGSLKKQSPKVT 900
Db 1331 RETSIEMESELDAQYLQNTFKVSKROSFALFNPNGNAEBECATFSAHSGSLKKQSPKVT 1390
QY 901 FECEQKEENOGKNEKNKPKVQTNITAGFPVVGOKDPVDNAKCSIKGGRFCLSSQFRG 960
Db 1391 FECEQKEENOGKNEKNKPKVQTNITAGFPVVGOKDPVDNAKCSIKGGRFCLSSQFRG 1450

Db 61 CPCKNDITKRSQESTFRFSQVVEELKIIICAFQDITGLEIYANSYNPAKKNENSPHKLKD 120
QY 121 EVSTIQSMGYNRAKRLQLQSEPNPSLQETSLSVOLSNLGTVRTLRKQIQOKTQSVYI 180
Db 121 EVSTIQSMGYNRAKRLQLQSEPNPSLQETSLSVOLSNLGTVRTLRKQIQOKTQSVYI 180
QY 181 ELGSDSSEDTVNNKATYCSVGQDQLLQITPQGTDRDEISLSDAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDTVNNKATYCSVGQDQLLQITPQGTDRDEISLSDAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAERHPEKXQSSVSNLHVEPCGINTHASSLOHENSLLLLTKDRMNVE 300
Db 241 PSNNDLNTTEKRAERHPEKXQSSVSNLHMEPCGINTHASSLOHENSLLLLTKDRMNVE 300
QY 301 KAEFCNKSQOGLARSOHNRWAGSKETCNDRTPTSTKKVLDLADPLCERKEWNKQKLPC 360
Db 301 KAEFCNKSQOGLARSOHNRWAGSKETCNDRTPTSTKKVLDLADPLCERKEWNKQKLPC 360
QY 361 SENPRDTEVPWITLNSSIQKVNEWFSRDELGSDSDHGESESNAKVADVLVDLNEVD 420
Db 361 SENPRDTEVPWITLNSSIQKVNEWFSRDELGSDSDHGESESNAKVADVLVDLNEVD 420
QY 421 EYSGSEKIDLLASDPHEALICKSERVHKSVSNSNIEDKIFGKTYRKASLPNLSHVTEN 480
Db 421 EYSGSEKIDLLASDPHEALICKSERVHKSVSNSNIEDKIFGKTYRKASLPNLSHVTEN 480
QY 481 LIIGAFVTEPOIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPMINOGTNOTE 540
Db 481 LIIGAFVTEPOIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPMINOGTNOTE 540
QY 541 QNGOVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI 600
Db 541 QNGOVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI 600
QY 601 HNSKAPKNLRLKRSSTRHIALELVVSRNLSPNCTELOIDSCSSSEETKKKYNQMPV 660
Db 601 HNSKAPKNLRLKRSSTRHIALELVVSRNLSPNCTELOIDSCSSSEETKKKYNQMPV 660
QY 661 RHRNRLQMEGKBPATGAKKSNKPNNEQTSKRHSDTFPELKLTNAPGSFTKCSNTSELKE 720
Db 661 RHRNRLQMEGKBPATGAKKSNKPNNEQTSKRHSDTFPELKLTNAPGSFTKCSNTSELKE 720
QY 721 FVNPSLPREEKEKLETVKVSNAEDPKDMLSGERVLOQTERSVESSISLVPDGTGTQ 780
Db 721 FVNPSLPREEKEKLETVKVSNAEDPKDMLSGERVLOQTERSVESSISLVPDGTGTQ 780
QY 781 ESTSLLEVSTLGRKAKTEPNKCVSOCAAFENPKGLIHCCSKDNNDTEGFKYPLGHEVNH 840
Db 781 ESTSLLEVSTLGRKAKTEPNKCVSOCAAFENPKGLIHCCSKDNNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMEESDELDAQYLQNTFKYSKRQSFALFSPNGNAEEECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMEESDELDAQYLQNTFKYSKRQSFALFSPNGNAEEECATFSAHSGSLKKQSPKVT 900
QY 901 FECEQKEENOGKNESNIKPVQTNITAGFPVVGQDKPVDNAKCSIKGGSFRCLSSQFRG 960
Db 901 FECEQKEENOGKNESNIKPVQTNITAGFPVVGQDKPVDNAKCSIKGGSFRCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNKLNEENFEHSMSPEREMGNENIP 1020
Db 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNKLNEENFEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNGPKL 1080
Db 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNGPKL 1080
QY 1081 NAMLRGLVQPEYIKOSLPGSNCKHPEIKKQEEYEEVVQTVNTDFSPYLISDNLEQPMGSS 1140
Db 1081 NAMLRGLVQPEYIKOSLPGSNCKHPEIKKQEEYEEVVQTVNTDFSPYLISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDDLLDGEIKEDTSPAENDIKESSAVFSKVQKGELSRSPFTHTHLAQ 1200

Db 1141 HASQVCSETPDDLLDGEIKEDTSPAENDIKESSAVFSKVQKGELSRSPFTHTHLAQ 1200
QY 1201 GYRRGAKKLESSEENLSSEDEELPCFQHLHFGKVNINIPSOSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSSEDEELPCFQHLHFGKVNINIPSOSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLSFSSQCSSELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLSFSSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKOMRQSESQGVGLSDKELVSDDEERGTCLEENNOEQSDNMGAAAGCSESESVSE 1380
Db 1321 SKOMRQSESQGVGLSDKELVSDDEERGTCLEENNOEQSDNMGAAAGCSESESVSE 1380
QY 1381 DCSGLSQSDILITQOQDRTMQHNLIKLOQEMAELEAVLEQHGQSPNSYPSIISDSALE 1440
Db 1381 DCSGLSQSDILITQOQDRTMQHNLIKLOQEMAELEAVLEQHGQSPNSYPSIISDSALE 1440
QY 1441 DLNRPQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPVGVERSPSK 1500
Db 1441 DLNRPQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPVGVERSPSK 1500
QY 1501 CPSLDDRWYMHSCSGSLQNRNYPQBELIKVVDVEEQOLEESGPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDDRWYMHSCSGSLQNRNYPQBELIKVVDVEEQOLEESGPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESGISLFSDDPSDDPSDRAPESARVGNIPSTTSALKVPQLKVAESAQSPAAAHTT 1620
Db 1561 TPYLESGISLFSDDPSDDPSDRAPESARVGNIPSTTSALKVPQLKVAESAQSPAAAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERNVKNRSMVVSGLTPEEFMLVYKFARKHHTLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERNVKNRSMVVSGLTPEEFMLVYKFARKHHTLTNLI 1680
QY 1681 TEETHVVMKTDFAEFCERTLKYFLGIAGGKVVVYFWTQTSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDFAEFCERTLKYFLGIAGGKVVVYFWTQTSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQQPKRARESODRKIFRGLEICCCYGPFTNMPDQLEMMVQLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQQPKRARESODRKIFRGLEICCCYGPFTNMPDQLEMMVQLCGASVVKELSSFTL 1800
QY 1801 GTGVHPVIVVQPDWATEDNGFHAIGQMCEAPVVTREWVLDVALYQCOELDTYLIQIPH 1860
Db 1801 GTGVHPVIVVQPDWATEDNGFHAIGQMCEAPVVTREWVLDVALYQCOELDTYLIQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 10
AAR81500
ID AAR81500 standard; Protein; 1863 AA.
XX
AC AAR81500;
XX AC
DT 02-OCT-1996 (first entry)
XX
DE BRCA1 mutant from sample sets YN98 and YN7.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 271
XX /note= "v271m"
XX
PN WO9605306-A2.
XX

PD 22-FEB-1996.
XX 11-AUG-1995; 95WO-US10202.
XX 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 94US-0409305.
PR 07-JUN-1995; 95US-0480784.
XX (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.
XX Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
PI WPI; 1996-139702/14.
DR N-PSDB; AAT17457.
XX New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
XX Claim 1; : 218pp; English.
XX AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
CC (see AAR81481 for wild type protein). These mutations can be used as
CC immunogens for antibody production. The mutant BRCA1 genes encoding
CC these sequences have at least 1 mutation or polymorphism in comparison
CC to the wild type cDNA (see AAT17438 for wild type). By detecting a
CC germline alteration in the wild type BRCA1 gene, a predisposition for
CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
CC isolated from a tissue sample from a subject has a probe, corresponding
CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an
CC allele-specific probe for a mutation of it), added to it. The conditions
CC allow for hybridisation of the probe to the mRNA, and any hybridisation
CC which occurs is detected. Alternatively the BRCA1 gene in the tissue
CC sample is isolated, and a shift in electrophoretic mobility of single
CC stranded DNA from the sample on a non-denaturing polyacrylamide gel
CC indicates a mutation. These methods of detection can also diagnose a
CC lesion neoplasia associated with the BRCA1 locus. The methods may be
CC used in gene therapy, protein replacement therapy and protein mimetics,
CC and may be used to screen for drugs in cancer therapy.
XX Sequence 1863 AA;
SQ Query Match 99.9%; Score 9639; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNKKGPSQ 60
DB 1 MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNKKGPSQ 60
QY 61 CPLCKNDITKRSLOESTRFSQVLELLKICAFQDGTGLEVANSYNFAKKNNSPEHLKD 120
DB 61 CPLCKNDITKRSLOESTRFSQVLELLKICAFQDGTGLEVANSYNFAKKNNSPEHLKD 120
QY 121 EYVSIQSMGYRNRAKRLQSEPNFSLOETSLSVLSNLGTVRTLTKRQIQPOKTSYVI 180
DB 121 EYVSIQSMGYRNRAKRLQSEPNFSLOETSLSVLSNLGTVRTLTKRQIQPOKTSYVI 180
QY 181 ELGSDSSEDTVNKATYCSVGDQELLIQITPQGTTRDEISLSAKKAACEFSETDVTNTEHQ 240
DB 181 ELGSDSSEDTVNKATYCSVGDQELLIQITPQGTTRDEISLSAKKAACEFSETDVTNTEHQ 240
QY 241 PSNNDLNTEKRAERHPKYGSSVSNLHVEPCGTNTNTHASSLQHNSSLLLTDRMNV 300
DB 241 PSNNDLNTEKRAERHPKYGSSVSNLHVEPCGTNTNTHASSLQHNSSLLLTDRMNV 300

QY 301 KAEFCNKSQGLARSOHRWAGSKETCNDRTPTSTKVKVDLNADPLCERKWNKOKLPC 360
DB 301 KAEFCNKSQGLARSOHRWAGSKETCNDRTPTSTKVKVDLNADPLCERKWNKOKLPC 360
QY 361 SENPRDTEVPWITLNSSTQKVNEWFSRDELLGSDSDSHGSESNKAVADVLDVLYNEVD 420
DB 361 SENPRDTEVPWITLNSSTQKVNEWFSRDELLGSDSDSHGSESNKAVADVLDVLYNEVD 420
QY 421 EYSGSSEKIDILLASDPHEALICKSERVHKS SVESNIEDKIFGKTYRKKASLNLSHV TEN 480
DB 421 EYSGSSEKIDILLASDPHEALICKSERVHKS SVESNIEDKIFGKTYRKKASLNLSHV TEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPDEFKKAADLAVOKTPEMINOQNTQTE 540
DB 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPDEFKKAADLAVOKTPEMINOQNTQTE 540
QY 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPPISSISNMELELNI 600
DB 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPPISSISNMELELNI 600
QY 601 HNSKAPKNRLRRKSSRHHIHALELVVSRNLSPPNCTELQIDSCSSSEBEIKKKYNQMPV 660
DB 601 HNSKAPKNRLRRKSSRHHIHALELVVSRNLSPPNCTELQIDSCSSSEBEIKKKYNQMPV 660
QY 661 RHSRNLQMEGKEPATGAKSKNKPNEQTSKRHSDSTEPELKLNTAPGSPFKCSNTSELKE 720
DB 661 RHSRNLQMEGKEPATGAKSKNKPNEQTSKRHSDSTEPELKLNTAPGSPFKCSNTSELKE 720
QY 721 FVNPSPREEKEKLETVKVSNNADDPKDLMSGERVLTQTERSVESSISLVPGTDYGTQ 780
DB 721 FVNPSPREEKEKLETVKVSNNADDPKDLMSGERVLTQTERSVESSISLVPGTDYGTQ 780
QY 781 ESISLLEVTSLGKAKTEPNKCVSQCAAFENPKGLIHGSKDNRNDETEGFKYPLGHEVNH 840
DB 781 ESISLLEVTSLGKAKTEPNKCVSQCAAFENPKGLIHGSKDNRNDETEGFKYPLGHEVNH 840
QY 841 RETSTEMESELDAQYLQNTFKVSKRQSFALFSPNPGNAEEECATFSAHSGSLKKQSPKVT 900
DB 841 RETSTEMESELDAQYLQNTFKVSKRQSFALFSPNPGNAEEECATFSAHSGSLKKQSPKVT 900
QY 901 FECQKEENQKNSNIKPQVTNITAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRG 960
DB 901 FECQKEENQKNSNIKPQVTNITAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKKNLLEENFEHSHMSPEREMGNENIP 1020
DB 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKKNLLEENFEHSHMSPEREMGNENIP 1020
QY 1021 STVSTISRNNIRENVFKEASSNINEVGSSTNEVGSSINEIGSSDENTQAEELGRNRPKL 1080
DB 1021 STVSTISRNNIRENVFKEASSNINEVGSSTNEVGSSINEIGSSDENTQAEELGRNRPKL 1080
QY 1081 NAMRLGVLPQEVYKQSLPGSNCKHPETKQIEYEVWQVTNTDFSPYLLISDLEQPMGSS 1140
DB 1081 NAMRLGVLPQEVYKQSLPGSNCKHPETKQIEYEVWQVTNTDFSPYLLISDLEQPMGSS 1140
QY 1141 HASQVCSPTPDDLDGGEIKEDTSFAENDIKESSAVFSKSVQKGELSRSPSFTHHLAQ 1200
DB 1141 HASQVCSPTPDDLDGGEIKEDTSFAENDIKESSAVFSKSVQKGELSRSPSFTHHLAQ 1200
QY 1201 GYRGAKKLESSEENLSEDEELPCFQHLFGKVNINIFSQSTRHSTVATECLSKNTEENL 1260
DB 1201 GYRGAKKLESSEENLSEDEELPCFQHLFGKVNINIFSQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDSCNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
DB 1261 LSLKNSLNDSCNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHQSOGVGLSKDELVDDEERTGTLEENNOEBSMDSNLGEAASCESETSVSE 1380
DB 1321 SKQMRHQSOGVGLSKDELVDDEERTGTLEENNOEBSMDSNLGEAASCESETSVSE 1380
QY 1381 DCSGLSSOSDILTTTQORDTMQHNLI KLOQEMAELEAVLEQHSQSPSNSYPSIISDSSALE 1440

Db 1381 DCSGLSQSDILTOQDTHQHNLIKLQEMAELEAVLEHGSGQPSNSYPSIISDSSALE 1440
QY 1441 DLNRPQSTSEKAVLTSOKSSEYPISONPEGLSADKFEVSADSTSKNKEPGVERSPSK 1500
Db 1441 DLNRPQSTSEKAVLTSOKSSEYPISONPEGLSADKFEVSADSTSKNKEPGVERSPSK 1500
QY 1501 CPSLDRWYMHSCSGSLQNRNYPQBELIKVVDVEEQOLEESGPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDRWYMHSCSGSLQNRNYPQBELIKVVDVEEQOLEESGPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESGISLFSDDPSDDPSDRAPESARVGNTPSTSAKVPOLKVAESAQSPAHAHTT 1620
Db 1561 TPYLESGISLFSDDPSDDPSDRAPESARVGNTPSTSAKVPOLKVAESAQSPAHAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFARKKHITNLNI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFARKKHITNLNI 1680
QY 1681 TEETHVWVKTDAEFCERTLKYFLGIAGGKWVYVFWVTQTSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVWVKTDAEFCERTLKYFLGIAGGKWVYVFWVTQTSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHOGPKRARESQRKIFRGLGIECCYGPFTNMPTDQLEWVQLCGASVVKELSSFTL 1800
Db 1741 VNGRNHOGPKRARESQRKIFRGLGIECCYGPFTNMPTDQLEWVQLCGASVVKELSSFTL 1800
QY 1801 GTGVHPVIVVQPDWATEDNGFHAIGOMCEAPVVTREWWLDSVALYOCQELDTYLIPQIPH 1860
Db 1801 GTGVHPVIVVQPDWATEDNGFHAIGOMCEAPVVTREWWLDSVALYOCQELDTYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 11
AAR81522
ID AAR81522 standard; Protein; 1863 AA.
AC AAR81522;
XX
DT 02-OCT-1996 (first entry)
XX
DE BRCA1 mutant from sample set MSK12871.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX
OS Homo sapiens.
XX
FH Key
FT Misc-difference 1219
FT /note= "E1219D"
XX
PN W09605306-A2.
XX
XX 22-FEB-1996.
XX
PF 11-AUG-1995; 95WO-US10202.
XX
PR 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0480784.
XX
XX (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.

XX Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
XX WPI: 1996-139702/14.
DR N-PSDB; AAT17479.
XX
PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
XX
PS Claim 1: ; 218pp; English.
XX
CC AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
CC (see AAR81481 for wild type protein). These mutations can be used as
CC immunogens for antibody production. The mutant BRCA1 genes encoding
CC these sequences have at least 1 mutation or polymorphism in comparison
CC to the wild type cDNA (see AAT17438 for wild type). By detecting a
CC germline alteration in the wild type BRCA1 gene, a predisposition for
CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
CC isolated from a tissue sample from a subject has a probe, corresponding
CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an
CC allele-specific probe for a mutation of it), added to it. The conditions
CC allow for hybridisation of the probe to the mRNA, and any hybridisation
CC which occurs is detected. Alternatively the BRCA1 gene in the tissue
CC sample is isolated, and a shift in electrophoretic mobility of single
CC stranded DNA from the sample on a non-denaturing polyacrylamide gel
CC indicates a mutation. These methods of detection can also diagnose a
CC lesion neoplasia associated with the BRCA1 locus. The methods may be
CC used in gene therapy, protein replacement therapy and protein mimetics,
CC and may be used to screen for drugs in cancer therapy.
XX
XX Sequence 1863 AA;
SQ
Query Match 99.9%; Score 9639; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEEVQNVINAMOKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKGPQSQ 60
Db 1 MDLSALRVEEVQNVINAMOKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKGPQSQ 60
QY 61 CPLCKNDITKRSLOESTRFSQVVEELLKIICAFQDGLTGLYANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVVEELLKIICAFQDGLTGLYANSYNFAKKENNSPEHLKD 120
QY 121 EYSIIQSMGYRNRRAKLLQSEPNPSLQFTSLSVQSLNLTGVTTLTKTORIQPKTSVYI 180
Db 121 EYSIIQSMGYRNRRAKLLQSEPNPSLQFTSLSVQSLNLTGVTTLTKTORIQPKTSVYI 180
QY 181 ELGSDSSEDTVNKATYCSVGDQELLQITPQGTREISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDTVNKATYCSVGDQELLQITPQGTREISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVEPCGNTNTHASSLQHENSILLTKDRMNYE 300
Db 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVEPCGNTNTHASSLQHENSILLTKDRMNYE 300
QY 301 KAEFCNKSKQPLARSQHNRWAGSKETCNDRRTPSTPEKKVDLNLADPLCERKENWKKLPK 360
Db 301 KAEFCNKSKQPLARSQHNRWAGSKETCNDRRTPSTPEKKVDLNLADPLCERKENWKKLPK 360
QY 361 SENPRDTEVPWITLNSSTOKYNEWFSRDELLGSDSDHGESESNKAVADVLDVLENYD 420
Db 361 SENPRDTEVPWITLNSSTOKYNEWFSRDELLGSDSDHGESESNKAVADVLDVLENYD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTE 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTE 480
QY 481 LIIGAFVTEPQIIQERPLTNKLKRRRP7SGLHPEDFIKKADLAVQKTPPEMINQGTNOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLKRRRP7SGLHPEDFIKKADLAVQKTPPEMINQGTNOTE 540

Qy 541 QNGQVNIITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSISNMELELNI 600
Dy 541 QNGQVNIITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSISNMELELNI 600
Qy 601 HNSKAPKNRLRRKSTRIHIALELWVRNLSPPNCTELOQDSCSSSEIEIKKKYNQMPV 660
Dy 601 HNSKAPKNRLRRKSTRIHIALELWVRNLSPPNCTELOQDSCSSSEIEIKKKYNQMPV 660
Qy 661 RISRNLQMEGKEPATGAKSKNPNQOTSKRHDSDTFPELKLITNAPGSFTKCSNTSELKE 720
Dy 661 RISRNLQMEGKEPATGAKSKNPNQOTSKRHDSDTFPELKLITNAPGSFTKCSNTSELKE 720
Qy 721 FVNPSLPREEKEKLETWKVSNNAEDPKDMLSGSERVLQTERSVESSISILVPGTDYGTQ 780
Dy 721 FVNPSLPREEKEKLETWKVSNNAEDPKDMLSGSERVLQTERSVESSISILVPGTDYGTQ 780
Qy 781 ESTLSLEVSTLGAKTEPNKCVSOCAAFENPKGLIHGCKDNRNDTEGFKYPLGHEVNHS 840
Dy 781 ESTLSLEVSTLGAKTEPNKCVSOCAAFENPKGLIHGCKDNRNDTEGFKYPLGHEVNHS 840
Qy 841 RETSTEMESELDAOYLONTFKVSKRQSFALFSPNPGNAEEBCATFSAHSGSLKKQSPKVT 900
Dy 841 RETSTEMESELDAOYLONTFKVSKRQSFALFSPNPGNAEEBCATFSAHSGSLKKQSPKVT 900
Qy 901 FECEQKEENQKNESNKPVTQVNTITAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRG 960
Dy 901 FECEQKEENQKNESNKPVTQVNTITAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRG 960
Qy 961 NETGLITPNKHGLLQNPRIPLPIKSFVTKCKKKNLLENFEEHSMSPEREMGNENIP 1020
Dy 961 NETGLITPNKHGLLQNPRIPLPIKSFVTKCKKKNLLENFEEHSMSPEREMGNENIP 1020
Qy 1021 SVVSTISRNNIRENVFKKASSNINEVGSSSTNEIGSSDENIQAEELGRNRPKL 1080
Dy 1021 SVVSTISRNNIRENVFKKASSNINEVGSSSTNEIGSSDENIQAEELGRNRPKL 1080
Qy 1081 NAMRLGLVLQPEVYKQSLPGSNCKHPEIKQEYEVVQTVNTDFSPYLISDNLEQPMGSS 1140
Dy 1081 NAMRLGLVLQPEVYKQSLPGSNCKHPEIKQEYEVVQTVNTDFSPYLISDNLEQPMGSS 1140
Qy 1141 HASQVCESTPDDLLDGEIKEDTSAENDIKESSAVFSKSVQKELSRSPFTHLQAQ 1200
Dy 1141 HASQVCESTPDDLLDGEIKEDTSAENDIKESSAVFSKSVQKELSRSPFTHLQAQ 1200
Qy 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFLGKYNNIPSQSTRSTVATECLSKNTEENL 1260
Dy 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFLGKYNNIPSQSTRSTVATECLSKNTEENL 1260
Qy 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSQCSLEDELDTANTWTQDPFLIGS 1320
Dy 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSQCSLEDELDTANTWTQDPFLIGS 1320
Qy 1321 SKOMRHQESQGVGLSDKELVSDDEERTGTLENNQBEQSMDSNLGRAASCESETSVSE 1380
Dy 1321 SKOMRHQESQGVGLSDKELVSDDEERTGTLENNQBEQSMDSNLGRAASCESETSVSE 1380
Qy 1381 DCSGLSSOSDILTTQOVRTMOHNLKLOEMAELVLEQHSQPSNSYPSIISDSSALE 1440
Dy 1381 DCSGLSSOSDILTTQOVRTMOHNLKLOEMAELVLEQHSQPSNSYPSIISDSSALE 1440
Qy 1441 DLRNPEQSTSEKAVLTSQKSSEYPISQNPEGLSADKFEVSADSSTSKNKEPVERSSPSK 1500
Dy 1441 DLRNPEQSTSEKAVLTSQKSSEYPISQNPEGLSADKFEVSADSSTSKNKEPVERSSPSK 1500
Qy 1501 CPSLDRYMHSCSSGLQNRYPSQBELIKVDVBEQOLESQPHDLTETSYLPRQOLEG 1560
Dy 1501 CPSLDRYMHSCSSGLQNRYPSQBELIKVDVBEQOLESQPHDLTETSYLPRQOLEG 1560
Qy 1561 TPYLESGISLSDPSPEDPESDRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHHT 1620
Dy 1561 TPYLESGISLSDPSPEDPESDRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHHT 1620

Qy 1621 DTAGYNAMESVSREKPELTASTERTVNRKMSVMVSGLTPPEEFMLVYKFARKHHITLTNLI 1680
Dy 1621 DTAGYNAMESVSREKPELTASTERTVNRKMSVMVSGLTPPEEFMLVYKFARKHHITLTNLI 1680
Qy 1681 TEETHVVMKTDADFVCERTLKIFYLGIAGGKVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
Dy 1681 TEETHVVMKTDADFVCERTLKIFYLGIAGGKVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
Qy 1741 VNGRNHOGPKRARESODRKIFRGLIEICCYGPTNNPTDQLEMMVOLCGASVVKELSSFTL 1800
Dy 1741 VNGRNHOGPKRARESODRKIFRGLIEICCYGPTNNPTDQLEMMVOLCGASVVKELSSFTL 1800
Qy 1801 GTGVHPPIVVQPDAWTEDNGFHAIGQMCAPVVTREWVLDVALYQCQELDTYLLIPQIPH 1860
Dy 1801 GTGVHPPIVVQPDAWTEDNGFHAIGQMCAPVVTREWVLDVALYQCQELDTYLLIPQIPH 1860
Qy 1861 SHY 1863
Dy 1861 SHY 1863
RESULT 12
AAR81536
ID AAR81536 standard; Protein; 1863 AA.
XX AAR81536;
XX AC AC
XX 02-OCT-1996 (first entry)
DT 02-OCT-1996 (first entry)
XX BRCA1 mutant from PM07.
DE
XX
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1183
FT /note= "K1183R"
XX
PN WO9605306-A2.
XX
XX 22-FEB-1996.
XX
XX 11-AUG-1995; 95WO-US10202.
XX
XX 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 12-AUG-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0480784.
XX
XX (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.
XX
PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
XX
DR WPI; 1996-139702/14.
DR N-PSDB; AAT17494.
XX
XX New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
XX
XX Claim 1; : 218pp; English.
PS
XX
CC AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
CC (see AAR81481 for wild type protein). These mutations can be used as

CC immunogens for antibody production. The mutant BRCA1 genes encoding
CC these sequences have at least 1 mutation or polymorphism in comparison
CC to the wild type cDNA (see AAT17438 for wild type). By detecting a
CC germline alteration in the wild type BRCA1 gene, a predisposition for
CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
CC isolated from a tissue sample from a subject has a probe, corresponding
CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an
CC allele-specific probe for a mutation of it), added to it. The conditions
CC allow for hybridisation of the probe to the mRNA, and any hybridisation
CC which occurs is detected. Alternatively the BRCA1 gene in the tissue
CC sample is isolated, and a shift in electrophoretic mobility of single
CC stranded DNA from the sample on a non-denaturing polyacrylamide gel
CC indicates a mutation. These methods of detection can also diagnose a
CC lesion neoplasia associated with the BRCA1 locus. The methods may be
CC used in gene therapy, protein replacement therapy and protein mimetics,
CC and may be used to screen for drugs in cancer therapy.

XX
SQ Sequence 1863 AA;

Query Match 99.9%; Score 9639; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVVEQVNVINAMQILECPICLELIKPEVSTKCDHIFCKFCMLKLLNKKGPSQ 60
DB 1 MDLSALRVVEQVNVINAMQILECPICLELIKPEVSTKCDHIFCKFCMLKLLNKKGPSQ 60

QY 61 CPICKNDITKRSLOESTRFSQVLEELKIIICAFOLDTGLEVANSYNFAKKENNSPEHLKD 120
DB 61 CPICKNDITKRSLOESTRFSQVLEELKIIICAFOLDTGLEVANSYNFAKKENNSPEHLKD 120

QY 121 EVSIIOGMGYNNRAKRLQLOSEPENPSLOETSLSVQLSNLGTVRTLRKQIQPKTSVYI 180
DB 121 EVSIIOGMGYNNRAKRLQLOSEPENPSLOETSLSVQLSNLGTVRTLRKQIQPKTSVYI 180

QY 181 ELGSDSSEDIVNATKATYCSVGDEQLLQITPQGTREISLDSAKKAACSEFSDVTNTSEHQ 240
DB 181 ELGSDSSEDIVNATKATYCSVGDEQLLQITPQGTREISLDSAKKAACSEFSDVTNTSEHQ 240

QY 241 PSNNDLNTTEKRAAERPEKYQSSVSNLHVPCGNTTHASSLQHSNLSLLTKDRMNV 300
DB 241 PSNNDLNTTEKRAAERPEKYQSSVSNLHVPCGNTTHASSLQHSNLSLLTKDRMNV 300

QY 301 KAEPCKNSKOPGLARSQHNWAGSKETCNDRRTPSTPEKKVDLNADPLCKERKNWKKLPC 360
DB 301 KAEPCKNSKOPGLARSQHNWAGSKETCNDRRTPSTPEKKVDLNADPLCKERKNWKKLPC 360

QY 361 SENPRDTEVPWITLNSSIQKVNEWFSRDELGSDSDSHGSESNAKAVADVLVNEVD 420
DB 361 SENPRDTEVPWITLNSSIQKVNEWFSRDELGSDSDSHGSESNAKAVADVLVNEVD 420

QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKVESNIEDKIFGKTYRKASLPNLSHVTEN 480
DB 421 EYSGSSEKIDLLASDPHEALICKSERVHSKVESNIEDKIFGKTYRKASLPNLSHVTEN 480

QY 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPDEFIKKADLAVQKTPEMINQGTNOTE 540
DB 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPDEFIKKADLAVQKTPEMINQGTNOTE 540

QY 541 QNGOVNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNWELENI 600
DB 541 QNGOVNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNWELENI 600

QY 601 HNSKAPKKNLRKRSSTRHIALVELVYSRNLSPNCPNTELOIDSCSSSEEEKKKKYNOMPV 660
DB 601 HNSKAPKKNLRKRSSTRHIALVELVYSRNLSPNCPNTELOIDSCSSSEEEKKKKYNOMPV 660

QY 661 RHSNRLQMECKEPATGAKKSNKPNNEQTSKRHSDSTFPPELKLTNAPGSFTKCSNTSELKE 720
DB 661 RHSNRLQMECKEPATGAKKSNKPNNEQTSKRHSDSTFPPELKLTNAPGSFTKCSNTSELKE 720

QY 721 FVNPSPREEKEKLETVKVSNNNAEDPKDMLSGERVLOTERSVESSSISLVPGTGYGTQ 780
DB 721 FVNPSPREEKEKLETVKVSNNNAEDPKDMLSGERVLOTERSVESSSISLVPGTGYGTQ 780

DB 721 FVNPSPREEKEKLETVKVSNNNAEDPKDMLSGERVLOTERSVESSSISLVPGTGYGTQ 780
QY 781 ESTISLLEVTILGRAKTEPNKCVSOCAAFENPKGLIHGCSKDNRRNDTEGFKYPLGHEVNH 840
DB 781 ESTISLLEVTILGRAKTEPNKCVSOCAAFENPKGLIHGCSKDNRRNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFNSNPGNAEEECATFSAHSGSLKKQSPKVT 900
DB 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFNSNPGNAEEECATFSAHSGSLKKQSPKVT 900
QY 901 FECEQKEENOGKKNESNITKPVQTNITAGFPVVGOKDPVDNAKCSIKGSRFCLSSQFRG 960
DB 901 FECEQKEENOGKKNESNITKPVQTNITAGFPVVGOKDPVDNAKCSIKGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLONPYRIPPLPIKSFVKTKCKNLEENFEHSMSPEREMGNENIP 1020
DB 961 NETGLITPNKHGLLONPYRIPPLPIKSFVKTKCKNLEENFEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNNIRENVFKEAASSNINEVGSSTNEVGSSEINIGSSDENIQAEIAGLRNRPGL 1080
DB 1021 STVSTISRNNIRENVFKEAASSNINEVGSSTNEVGSSEINIGSSDENIQAEIAGLRNRPGL 1080
QY 1081 NAMLRGLVLOPEVYKQSLPGSNCKHPKIKQOEYEEVQTVNTDFSPYLIISDNLEQPMGSS 1140
DB 1081 NAMLRGLVLOPEVYKQSLPGSNCKHPKIKQOEYEEVQTVNTDFSPYLIISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDLLDDGEIKEIDTSFAENDIKESAVFSKSVQKGEISKSPSPFTHHLAQ 1200
DB 1141 HASQVCSETPDLLDDGEIKEIDTSFAENDIKESAVFSKSVQKGEISKSPSPFTHHLAQ 1200
QY 1201 GYRRGAKKLESSENLSSEDEELPCFOHLLFGKVNIPSPQSTRHSTVATECLSKNTEENL 1260
DB 1201 GYRRGAKKLESSENLSSEDEELPCFOHLLFGKVNIPSPQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDSCNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
DB 1261 LSLKNSLNDSCNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKOMRHQSESGVGLSKDELVSDDERGTGLEENNQEQQSMDSNLGEAASGCESETSVSE 1380
DB 1321 SKOMRHQSESGVGLSKDELVSDDERGTGLEENNQEQQSMDSNLGEAASGCESETSVSE 1380
QY 1381 DCSGLSSQSDILTTQQRDTHQNLIKLOQEMAELEAVLEQHGQSPNSYPSIISDSSALE 1440
DB 1381 DCSGLSSQSDILTTQQRDTHQNLIKLOQEMAELEAVLEQHGQSPNSYPSIISDSSALE 1440
QY 1441 DLNRPEOSTSEKAVLTISQKSSEYPISONPEGLSADKEFEVSADSTSKNKEPGEVERSPSK 1500
DB 1441 DLNRPEOSTSEKAVLTISQKSSEYPISONPEGLSADKEFEVSADSTSKNKEPGEVERSPSK 1500
QY 1501 CPSLDDRWYMHSCSGSLQNRNYPQEBELIKVVDVEEQOLEESGPHDLTETSYLPRDLEG 1560
DB 1501 CPSLDDRWYMHSCSGSLQNRNYPQEBELIKVVDVEEQOLEESGPHDLTETSYLPRDLEG 1560
QY 1561 TPYLESGISLFSDDPSDDPSDRAPESARVGNIPSSTSALKVPQLKVAESAQSPAAAHFT 1620
DB 1561 TPYLESGISLFSDDPSDDPSDRAPESARVGNIPSSTSALKVPQLKVAESAQSPAAAHFT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERNVKNRMVMVSGLTPEEFMLVYKFARKHHTITLNL 1680
DB 1621 DTAGYNAMESVSREKPELTASTERNVKNRMVMVSGLTPEEFMLVYKFARKHHTITLNL 1680
QY 1681 TEETHVMKTDFAEVCERTLKVFGLGAGKVVYVFWVTQSTIKERKMLNEHDFEVRGDV 1740
DB 1681 TEETHVMKTDFAEVCERTLKVFGLGAGKVVYVFWVTQSTIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHOGPKRARESQRDKIFRGLTCCYGPFTNMPTDQLEWMVQLGASVVKSELSSFTL 1800
DB 1741 VNGRNHOGPKRARESQRDKIFRGLTCCYGPFTNMPTDQLEWMVQLGASVVKSELSSFTL 1800
QY 1801 GTGVHPITVVVQPDWATEDNGFHAIGQMCAPVVTREWLDSVALYQCOELDTYLIPQIPH 1860
DB 1801 GTGVHPITVVVQPDWATEDNGFHAIGQMCAPVVTREWLDSVALYQCOELDTYLIPQIPH 1860

QY 1861 SHY 1863
III
Db 1861 SHY 1863

RESULT 13
AAR81540
ID AAR81540 standard; Protein; 1863 AA.
XX AAR81540;
XX
DT 02-OCT-1996 (first entry)
XX
DE BRCA1 mutant from PM20.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1040 /note= "S1040N"
FT
XX
PN WO96050306-A2.
XX
PD 22-FEB-1996.
XX
PF 11-AUG-1995; 95WO-US10202.
XX
PR 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0480784.
XX
PA (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHOL.
XX
PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
XX
DR WPI: 1996-139702/14.
DR N-PSDB; AAT17502.
XX
PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
XX
PS Claim 1; ; 218pp; English.

AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein encoded by the human breast and ovarian cancer predisposing gene (BRCA1) (see AAR81481 for wild type protein). These mutations can be used as immunogens for antibody production. The mutant BRCA1 genes encoding these sequences have at least 1 mutation or polymorphism in comparison to the wild type cDNA (see AAT17438 for wild type). By detecting a germline alteration in the wild type BRCA1 gene, a predisposition for breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA isolated from a tissue sample from a subject has a probe, corresponding to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an allele-specific probe for a mutation of it), added to it. The conditions allow for hybridisation of the probe to the mRNA, and any hybridisation which occurs is detected. Alternatively the BRCA1 gene in the tissue sample is isolated, and a shift in electrophoretic mobility of single stranded DNA from the sample on a non-denaturing polyacrylamide gel indicates a mutation. These methods of detection can also diagnose a lesion neoplasia associated with the BRCA1 locus. The methods may be used in gene therapy, protein replacement therapy and protein mimetics,

CC and may be used to screen for drugs in cancer therapy.
XX
SQ Sequence 1863 AA;
Query Match 99.9%; Score 9639; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEEVQNVINAMOKIIECPICLLEIKPEVSTKCDHIFCKFCMLKLLNKKGPSQ 60
Db 1 MDLSALRVEEVQNVINAMOKIIECPICLLEIKPEVSTKCDHIFCKFCMLKLLNKKGPSQ 60
QY 61 CPLCKNDITKRSLOESTRFSQVLELLKTIICAFQDITGLYANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVLELLKTIICAFQDITGLYANSYNFAKKENNSPEHLKD 120
QY 121 EYSITQSMGYNRKRLLOSEPNSLOETSLVQLSNLGTVRTLTQRIOQPKTSVYI 180
Db 121 EYSITQSMGYNRKRLLOSEPNSLOETSLVQLSNLGTVRTLTQRIOQPKTSVYI 180
QY 181 ELGSDSEDTYNKATYCSVGDOELLQITPQGTREISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSEDTYNKATYCSVGDOELLQITPQGTREISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNDLNTTEKRAERHPEKYQGSVSNLHVPCGTNTHASSLQHENSSLLLTKDRMNV 300
Db 241 PSNDLNTTEKRAERHPEKYQGSVSNLHVPCGTNTHASSLQHENSSLLLTKDRMNV 300
QY 301 KAEFCNKSQPLARSOHNRWAGSKETCNDRTPTSTKKYVDLNADPLCERKENWKKLPC 360
Db 301 KAEFCNKSQPLARSOHNRWAGSKETCNDRTPTSTKKYVDLNADPLCERKENWKKLPC 360
QY 361 SENPRDTEVPWITLNSSTOKVNEFSDLLGSDSDHGESENAKVADVDLVNEVD 420
Db 361 SENPRDTEVPWITLNSSTOKVNEFSDLLGSDSDHGESENAKVADVDLVNEVD 420
QY 421 EYSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHVTEN 480
Db 421 EYSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHVTEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFTKKADLAVOKTPEMINQGTNQE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFTKKADLAVOKTPEMINQGTNQE 540
QY 541 QNGQVMNITNSGHNKTKGDSIQNEKNPNPIESLSEKSAFKTKAEPISISSINMWELELNI 600
Db 541 QNGQVMNITNSGHNKTKGDSIQNEKNPNPIESLSEKSAFKTKAEPISISSINMWELELNI 600
QY 601 HNSKAPKNRLRRKSSTRHIIHALELVSRNLSPPNCTELQIDSCSSSEIEIKKKYQMPV 660
Db 601 HNSKAPKNRLRRKSSTRHIIHALELVSRNLSPPNCTELQIDSCSSSEIEIKKKYQMPV 660
QY 661 RHRNLQJMEGKEPATGAKSNKNEQTSKRHSDSTFPPELKLTNAPGSGTSCNSELKE 720
Db 661 RHRNLQJMEGKEPATGAKSNKNEQTSKRHSDSTFPPELKLTNAPGSGTSCNSELKE 720
QY 721 FVNPSLPEEKEELETYKVSNNADPKDMLSGERVLTQTSRVSSSISLVPGTDYGTQ 780
Db 721 FVNPSLPEEKEELETYKVSNNADPKDMLSGERVLTQTSRVSSSISLVPGTDYGTQ 780
QY 781 ESIISLLEVTSLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
Db 781 ESIISLLEVTSLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLQNTFKVSRQSFALFSPNGNAEEECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSRQSFALFSPNGNAEEECATFSAHSGSLKKQSPKVT 900
QY 901 FECQKEENQCKNESNIKPVQTVNITAGFPVVGQDKDPVDNAKCSIKGSGRCLSSQFRG 960
Db 901 FECQKEENQCKNESNIKPVQTVNITAGFPVVGQDKDPVDNAKCSIKGSGRCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLFPKISFVKTKCKKNLLLENFEEHSMSPERMGNIPI 1020


```

Db 961  |||||  NETGLITPNKHGLQNPYRIPPLFPKSFVKTKCKNLLNEENFEHSPERDMGNIP 1020
Qy 1021  |||||  STVSTIRNNIRENVFKEASSNINEVGSTNEVGSINEIGSSDENIOAELGRNGPKL 1080
Db 1021  |||||  STVSTIRNNIRENVFKEASSNINEVGSTNEVGSINEIGSSDENIOAELGRNGPKL 1080
Qy 1081  |||||  NAMLRGLVQLQPEYKQSLPGSNCKHPEIKKQEEYEVVQVNTDFSPYLLSDNLEQPMGSS 1140
Db 1081  |||||  NAMLRGLVQLQPEYKQSLPGSNCKHPEIKKQEEYEVVQVNTDFSPYLLSDNLEQPMGSS 1140
Qy 1141  |||||  HASQVCSETPDDLDDGETKEDTSFANDIKESSAVFSKSVQKGLSPSPPTHTHLLAQ 1200
Db 1141  |||||  HASQVCSETPDDLDDGETKEDTSFANDIKESSAVFSKSVQKGLSPSPPTHTHLLAQ 1200
Qy 1201  |||||  GYRRGAKKLESSBENLSSDEELPCFQHLLEGKVNIPQSTHRSTVATECLSKNTEENL 1260
Db 1201  |||||  GYRRGAKKLESSBENLSSDEELPCFQHLLEGKVNIPQSTHRSTVATECLSKNTEENL 1260
Qy 1261  |||||  LSLKNSLNCDSNOVILAKASQEHHLSEETKCSASLFSQCSLEEDLTANTQDPFLIGS 1320
Db 1261  |||||  LSLKNSLNCDSNOVILAKASQEHHLSEETKCSASLFSQCSLEEDLTANTQDPFLIGS 1320
Qy 1321  |||||  SKQMRHQSOGVGLSDKELVSDDERGTGLPENNOQEESMDSNLGEAASGCESETSVSE 1380
Db 1321  |||||  SKQMRHQSOGVGLSDKELVSDDERGTGLPENNOQEESMDSNLGEAASGCESETSVSE 1380
Qy 1381  |||||  DCSGLSSQSDILTTQORDTMOHNLKLOQEMAELEAVLEHQSGQSPNSVPSIISDSALE 1440
Db 1381  |||||  DCSGLSSQSDILTTQORDTMOHNLKLOQEMAELEAVLEHQSGQSPNSVPSIISDSALE 1440
Qy 1441  |||||  DLRNPEQSTSEKAVLTSQKSSEYPIQSONPEGLSADKFEVSADSTSKNKEPGVERSPSK 1500
Db 1441  |||||  DLRNPEQSTSEKAVLTSQKSSEYPIQSONPEGLSADKFEVSADSTSKNKEPGVERSPSK 1500
Qy 1501  |||||  CPSLDRWYMHSCGSLQNRNYPQBELIKVVDVEEQLEESGPHDLTSTYLPQDLEG 1560
Db 1501  |||||  CPSLDRWYMHSCGSLQNRNYPQBELIKVVDVEEQLEESGPHDLTSTYLPQDLEG 1560
Qy 1561  |||||  TPYLESGISLFDPSDDPSDEDRAPESARVGNIPSTSAKVPOLKVAESAQSPAAHHT 1620
Db 1561  |||||  TPYLESGISLFDPSDDPSDEDRAPESARVGNIPSTSAKVPOLKVAESAQSPAAHHT 1620
Qy 1621  |||||  DTAGYNAMESVSREKPELTASTERYNKRMVMVVSGLTPEEFMLVYKFARKHHITLNL 1680
Db 1621  |||||  DTAGYNAMESVSREKPELTASTERYNKRMVMVVSGLTPEEFMLVYKFARKHHITLNL 1680
Qy 1681  |||||  TEETHVVMKTDAEFVCERTLKYFLGIAGKVVVSYFWVTQTSIKERKMLNEHDFEVRGDV 1740
Db 1681  |||||  TEETHVVMKTDAEFVCERTLKYFLGIAGKVVVSYFWVTQTSIKERKMLNEHDFEVRGDV 1740
Qy 1741  |||||  VNGRNHOGPKRARESODRKIFRGLETCYGPFTNMPTDQLEWMVQJCGASVYKELSSFTL 1800
Db 1741  |||||  VNGRNHOGPKRARESODRKIFRGLETCYGPFTNMPTDQLEWMVQJCGASVYKELSSFTL 1800
Qy 1801  |||||  GTGVHPVVVVQPDWNTDNGFHAIGQMCAPVVTREWLWDSVALYOCQELDYLIPOIPH 1860
Db 1801  |||||  GTGVHPVVVVQPDWNTDNGFHAIGQMCAPVVTREWLWDSVALYOCQELDYLIPOIPH 1860
Qy 1861  |||||  SHY 1863
Db 1861  |||||  SHY 1863
```

RESULT 14

AAR81529

ID AAR81529 standard; Protein; 1863 AA.

XX

AC AAR81529;

XX

DT 02-OCT-1996 (first entry)

XX

DE BRCA1 mutant from sample set MSK9646.

```

XX KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FH Misc-difference 1628
FT FT /note= "M1628V"
XX XX
XX PN WO96050306-A2.
XX XX
XX PD 22-FEB-1996.
XX XX
XX PF 11-AUG-1995; 95WO-US10202.
XX XX
XX PR 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0480784.
XX XX
XX PA (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.
XX XX
XX PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
XX XX
XX DR WPI: 1996-139702/14.
DR N-PSDB; AAT17486.
XX XX
XX PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
XX PT gene - for diagnosis and therapy of human breast and ovarian cancer
XX PT and for diagnosing pre-disposition to these cancers
XX PS Claim 1; ; 218pp; English.
XX XX
XX CC AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
XX CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
XX CC (see AAR81481 for wild type protein). These mutations can be used as
XX CC immunogens for antibody production. The mutant BRCA1 genes encoding
XX CC these sequences have at least 1 mutation or polymorphism in comparison
XX CC to the wild type cDNA (see AAT17438 for wild type). By detecting a
XX CC germline alteration in the wild type BRCA1 gene, a predisposition for
XX CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
XX CC isolated from a tissue sample from a subject has a probe, corresponding
XX CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an
XX CC allele-specific probe for a mutation of it), added to it. The conditions
XX CC allow for hybridisation of the probe to the mRNA, and any hybridisation
XX CC which occurs is detected. Alternatively the BRCA1 gene in the tissue
XX CC sample is isolated, and a shift in electrophoretic mobility of single
XX CC stranded DNA from the sample on a non-denaturing polyacrylamide gel
XX CC indicates a mutation. These methods of detection can also diagnose a
XX CC lesion neoplasia associated with the BRCA1 locus. The methods may be
XX CC used in gene therapy, protein replacement therapy and protein mimetics,
XX CC and may be used to screen for drugs in cancer therapy.
XX SQ Sequence 1863 AA;
```

Query Match 99.9%; Score 9638; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MDLSALRVEEVQNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ 60

Db 1 MDLSALRVEEVQNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ 60

Qy 61 CPLCKNDITKRSIQESTRFSQVLELLKTKICAFQDGTGLEAYNSYFACKNNSPHLKD 120

Db 61 CPLCKNDITKRSIQESTRFSQVLELLKTKICAFQDGTGLEAYNSYFACKNNSPHLKD 120

QY 121 EYSIIQSMGYRNRKRLQSEPNPSLOETSLSVOLSNLGTVRTLRTKORIQPOKTSYVI 180
DB 121 EYSIIQSMGYRNRKRLQSEPNPSLOETSLSVOLSNLGTVRTLRTKORIQPOKTSYVI 180
QY 181 ELGSDSESDTVNKATYCSVGDQELQITPQGTTRDEISLDSAKAAECFSESDVTNTEHHQ 240
DB 181 ELGSDSESDTVNKATYCSVGDQELQITPQGTTRDEISLDSAKAAECFSESDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAEHPHEKYOGSSVSNLHVPCGTNTTHASSLOHENSLLLTTRDMNVE 300
DB 241 PSNNDLNTTEKRAAEHPHEKYOGSSVSNLHVPCGTNTTHASSLOHENSLLLTTRDMNVE 300
QY 301 KAEFCNKSQPGGLARSQHNRWAGSKETCNDRRTPSTTEKKVDLADPLCKERENKQKLP 360
DB 301 KAEFCNKSQPGGLARSQHNRWAGSKETCNDRRTPSTTEKKVDLADPLCKERENKQKLP 360
QY 361 SENPRTEDEVMITUNSSIQKYNWFPSRDELGLGSDSDHGDSESNNAKVAADVLDVNEVD 420
DB 361 SENPRTEDEVMITUNSSIQKYNWFPSRDELGLGSDSDHGDSESNNAKVAADVLDVNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHKS SVESNIEDKIFGKTYRKKASLPNLSHYTEN 480
DB 421 EYSGSSEKIDLLASDPHEALICKSERVHKS SVESNIEDKIFGKTYRKKASLPNLSHYTEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPEMINOGTNOTE 540
DB 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPEMINOGTNOTE 540
QY 541 QNGQVMNITNSGHEKNTKGDSTQNEKNPNPIESLEKESAFKTKAEPITSSSTSNMELELNI 600
DB 541 QNGQVMNITNSGHEKNTKGDSTQNEKNPNPIESLEKESAFKTKAEPITSSSTSNMELELNI 600
QY 601 HNSKAPKKNLRKSTRIHIALELVVRNLSPPNCTELOIDSCSSSEIEIKKKYQNPV 660
DB 601 HNSKAPKKNLRKSTRIHIALELVVRNLSPPNCTELOIDSCSSSEIEIKKKYQNPV 660
QY 661 RHSRLQLMEGREPATGAKSKNPNQETSKRHDSDTFPELKLITNAPGSFTKCSNTSELKE 720
DB 661 RHSRLQLMEGREPATGAKSKNPNQETSKRHDSDTFPELKLITNAPGSFTKCSNTSELKE 720
QY 721 FYNPSLPREEKEKLETYKVSNAEDDKDMLSGERVLTQERSVSSSISLVPGTDYGTQ 780
DB 721 FYNPSLPREEKEKLETYKVSNAEDDKDMLSGERVLTQERSVSSSISLVPGTDYGTQ 780
QY 781 ESISLLEVSTLGKATPEKNKCSQCAAFENPKGLTHGCSKDNRNDTEGKYPLGHEVNH 840
DB 781 ESISLLEVSTLGKATPEKNKCSQCAAFENPKGLTHGCSKDNRNDTEGKYPLGHEVNH 840
QY 841 RETSIEMEESLDAQYLQNTFKVSKRSQFALFSPGNAEEECATFSAHSGSLKKQSPKVT 900
DB 841 RETSIEMEESLDAQYLQNTFKVSKRSQFALFSPGNAEEECATFSAHSGSLKKQSPKVT 900
QY 901 FCEQKEENOGKNESNIKPQVTNITAGFPVVGOKDPVDNAKSIKGGSRFCLSSQFRG 960
DB 901 FCEQKEENOGKNESNIKPQVTNITAGFPVVGOKDPVDNAKSIKGGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGGLQNPYRIPPLPIKSFVTKCKKNLLEENFEHSMSPEREMGNENIP 1020
DB 961 NETGLITPNKHGGLQNPYRIPPLPIKSFVTKCKKNLLEENFEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNNIRENVFKASSNINEVSGSSNINEIGSSSDENITQABLGRNRPKL 1080
DB 1021 STVSTISRNNIRENVFKASSNINEVSGSSNINEIGSSSDENITQABLGRNRPKL 1080
QY 1081 NAMRLGLVQLPEVYKQSLPGSNCKHPETKQYEEVWQTVNTDFSPYLISDNLBQPMGSS 1140
DB 1081 NAMRLGLVQLPEVYKQSLPGSNCKHPETKQYEEVWQTVNTDFSPYLISDNLBQPMGSS 1140
QY 1141 HASQVCSETPDOLDGGEIKEDTFAENDIKESSAVFSKSVQKSGELSRSPSPFTHTHAQ 1200
DB 1141 HASQVCSETPDOLDGGEIKEDTFAENDIKESSAVFSKSVQKSGELSRSPSPFTHTHAQ 1200

QY 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFGKVNNIQSQTRHSTVATECLSKNTEENL 1260
DB 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFGKVNNIQSQTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLUNCNQVILAKASQEHHLSEETKCSASLFSQCSELEDLTANTNTQDPLIGS 1320
DB 1261 LSLKNSLUNCNQVILAKASQEHHLSEETKCSASLFSQCSELEDLTANTNTQDPLIGS 1320
QY 1321 SKQMRHOSGOGVGLSKELYSDDOERCTGLEENNOESQMSDNLGEAASCESETSVE 1380
DB 1321 SKQMRHOSGOGVGLSKELYSDDOERCTGLEENNOESQMSDNLGEAASCESETSVE 1380
QY 1381 DCSGLSSQSDILTTQQORDTMQHNLIKLOEMAELAVLEQHGQSQPSNSVPSTIISDSSALE 1440
DB 1381 DCSGLSSQSDILTTQQORDTMQHNLIKLOEMAELAVLEQHGQSQPSNSVPSTIISDSSALE 1440
QY 1441 DLNRPQSTSEKAVLTQSOKSEYPISONPEGLSADKFEVSADSSTSKNKEPQVERSSPSK 1500
DB 1441 DLNRPQSTSEKAVLTQSOKSEYPISONPEGLSADKFEVSADSSTSKNKEPQVERSSPSK 1500
QY 1501 CPSLDDRWYMHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPQDDLEG 1560
DB 1501 CPSLDDRWYMHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPQDDLEG 1560
QY 1561 TPYLESGLISLFDSDPESDRAPE SARVGNIPSSSTSALKVPQLKVAESAQSPAAAHTT 1620
DB 1561 TPYLESGLISLFDSDPESDRAPE SARVGNIPSSSTSALKVPQLKVAESAQSPAAAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVWSGLTPEEFMLVYKFAKHHITLNL 1680
DB 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVWSGLTPEEFMLVYKFAKHHITLNL 1680
QY 1681 TEETHVVMKTDAEFVCERTLYFLGIAGGKVVYSYFVWTSIKERKMLNEHDFEVRGDV 1740
DB 1681 TEETHVVMKTDAEFVCERTLYFLGIAGGKVVYSYFVWTSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHGPKRARESQRKIFRGLIICCYPFTNNPTDQLEMMVQLCGASVVKELSSFTL 1800
DB 1741 VNGRNHGPKRARESQRKIFRGLIICCYPFTNNPTDQLEMMVQLCGASVVKELSSFTL 1800
QY 1801 GTGVHPVIVVQPDWATEDNGFHAIGOMCEAPVVTREWVLDVSVALYQCQELDTYLIQIPH 1860
DB 1801 GTGVHPVIVVQPDWATEDNGFHAIGOMCEAPVVTREWVLDVSVALYQCQELDTYLIQIPH 1860
QY 1861 SHY 1863
DB 1861 SHY 1863
RESULT 15
AAR81532
ID AAR81532 standard; Protein; 1863 AA.
XX AAR81532;
DT 02-OCT-1996 (first entry)
XX AAR81532;
XX BRCA1 mutant from sample set MSK7542.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 1852
FT /note= "r1852s"
XX
XX W09605306-A2.
XX
XX 22-FEB-1996.
XX

Qy	1441	DLRNPQSTSEKAVLTSQKSSSEYPISQNPEGLSADKFEVSADSSSTSKNKEPGEVERSSPSK	1500
Db	1441	DLRNPQSTSEKAVLTSQKSSSEYPISQNPEGLSADKFEVSADSSSTSKNKEPGEVERSSPSK	1500
Qy	1501	CPSLDDRWMYHMSCGSLQNRNYPQSEELIKVVDVEEQOLEESGPHDITETSYLPRQDLEG	1560
Db	1501	CPSLDDRWMYHMSCGSLQNRNYPQSEELIKVVDVEEQOLEESGPHDITETSYLPRQDLEG	1560
Qy	1561	TPYLESIGLSLSDPSPEDSRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT	1620
Db	1561	TPYLESIGLSLSDPSPEDSRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT	1620
Qy	1621	DTAGYNAMESYSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFKARKHITLTNLI	1680
Db	1621	DTAGYNAMESYSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFKARKHITLTNLI	1680
Qy	1681	TEETHVWMKTDAEFCERTLKYFLGIAGGKWVSYFWVTQSIKERKMLNEHDFEVRGDV	1740
Db	1681	TEETHVWMKTDAEFCERTLKYFLGIAGGKWVSYFWVTQSIKERKMLNEHDFEVRGDV	1740
Qy	1741	VNGRNHQGPKRARESQRDKIFRGLBICCYGPTNMPDQLEWVMVLGCASVVKWLSFTL	1800
Db	1741	VNGRNHQGPKRARESQRDKIFRGLBICCYGPTNMPDQLEWVMVLGCASVVKWLSFTL	1800
Qy	1801	GTGVHPVVVQPDWANTEDNGFHAIGOMCEAPVPTREWVLDSVALYQCQELDLYLIPQIPH	1860
Db	1801	GTGVHPVVVQPDWANTEDNGFHAIGOMCEAPVPTREWVLDSVALYQCQELDLYLIPQIPH	1860
Qy	1861	SHY 1863	
Db	1861	SHY 1863	

Search completed: January 22, 2003, 16:46:04
Job time : 76 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:46:04 ; Search time 15.5 Seconds
(without alignments)
2425.332 Million cell updates/sec

Title: US-09-734-672-4

Perfect score: 9649

Sequence: 1 MDLSALRVEEVQNVINAMQK.....LYQCOELDTYLPQIPHSY 1863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	9649	100.0	1863	9 US-09-734-672-4	Sequence 4, Appli
2	9635	99.9	1863	9 US-09-734-672-2	Sequence 2, Appli
3	9635	99.9	1863	9 US-09-734-672-6	Sequence 6, Appli
4	357	3.7	2344	10 US-09-815-242-12713	Sequence 12713, A
5	341	3.5	6281	10 US-09-815-242-12996	Sequence 12996, A
6	328.5	3.4	2368	10 US-09-815-242-5635	Sequence 5635, Ap
7	328.5	3.4	2368	10 US-09-815-242-12389	Sequence 12389, A
8	314	3.3	1596	9 US-09-902-432-4	Sequence 4, Appli
9	312	3.2	1400	10 US-09-764-176-7	Sequence 7, Appli
10	312	3.2	2478	10 US-09-815-242-5816	Sequence 5816, Ap
11	312	3.2	2478	10 US-09-815-242-12967	Sequence 12967, A
12	308	3.2	2665	10 US-09-864-761-34248	Sequence 34248, A
13	306	3.2	3158	10 US-09-815-242-12611	Sequence 12611, A
14	299.5	3.1	2843	9 US-09-987-482-1	Sequence 1, Appli
15	298.5	3.1	2843	8 US-08-681-219-32	Sequence 32, Appli
16	286.5	3.0	1781	10 US-09-738-877-3	Sequence 3, Appli
17	285.5	3.0	2025	10 US-09-815-242-5703	Sequence 5703, Ap
18	282	2.9	50	10 US-09-998-667-15	Sequence 15, Appl
19	278	2.9	3256	10 US-09-919-172-98	Sequence 98, Appl

20	272	2.8	2437	10 US-09-815-242-5834	Sequence 5834, Ap
21	268	2.8	5795	10 US-09-815-242-12610	Sequence 12610, A
22	261	2.7	1285	10 US-09-982-091A-2	Sequence 2, Appli
23	258	2.7	1786	9 US-09-742-096-3	Sequence 3, Appli
24	255	2.6	2434	10 US-09-815-242-5835	Sequence 5835, Ap
25	253.5	2.6	1269	10 US-09-815-242-13113	Sequence 13113, A
26	251.5	2.6	1325	10 US-09-864-761-35612	Sequence 35612, A
27	251	2.6	2125	10 US-09-919-172-29	Sequence 29, Appl
28	250.5	2.6	1501	10 US-09-924-154-17	Sequence 17, Appl
29	250.5	2.6	2086	10 US-09-815-242-5639	Sequence 5639, Ap
30	250	2.6	2789	10 US-09-801-574-57	Sequence 57, Appl
31	248.5	2.6	2139	10 US-09-727-384-6	Sequence 6, Appli
32	248	2.6	1215	10 US-09-815-242-5908	Sequence 5908, Ap
33	241.5	2.5	1111	10 US-09-815-242-12955	Sequence 12955, A
34	235.5	2.4	1332	10 US-09-982-091A-4	Sequence 4, Appli
35	235	2.4	48	10 US-09-864-761-37757	Sequence 37757, A
36	234.5	2.4	1618	9 US-09-963-875-1	Sequence 1, Appli
37	234.5	2.4	1884	10 US-09-785-770A-17	Sequence 17, Appl
38	234.5	2.4	1907	10 US-09-785-770A-16	Sequence 16, Appl
39	232.5	2.4	2548	10 US-09-851-682A-1	Sequence 1, Appli
40	231.5	2.4	1346	9 US-09-902-432-2	Sequence 2, Appli
41	230.5	2.4	2308	12 US-10-000-954-2	Sequence 2, Appli
42	227.5	2.4	1086	10 US-09-924-154-15	Sequence 15, Appl
43	227.5	2.4	1597	9 US-09-832-292-35	Sequence 35, Appl
44	226.5	2.3	2785	10 US-09-801-574-8	Sequence 8, Appli
45	219.5	2.3	1478	10 US-09-801-368-52	Sequence 52, Appl

ALIGNMENTS

RESULT 1
US-09-734-672-4
; Sequence 4, Application US/09734672
; Publication No. US20020183268A1
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; Allen, Antonette C.
; Alvarez, Christopher P.
; Critz, Brenda S.
; Olson, Sheri J.
; Schelter, Denise B.
; Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; BRCA1 Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan Lewis & Bockius LLP
; STREET: 1111 Pennsylvania Ave., N.W.
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/734,672
; APPLICATION NUMBER: US/09/734,672
; FILING DATE: 03-Dec-2000
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/966,436
; FILING DATE: 07-No. US20020183268A1-97
; APPLICATION NUMBER: US 08/598,591
; FILING DATE: 12-Feb-96
; ATTORNEY/AGENT INFORMATION:
; NAME: Michael S. Tuscan
; REGISTRATION NUMBER: 43,210
; REFERENCE/DOCKET NUMBER: 44921-5055-02-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-739-3000
; TELEFAX: 202-739-3001

; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: No. US20020183268A1 Relevant
; TOPOLOGY: No. US20020183268A1 Relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCAL
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17G21
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-734-672-4

Query Match 100.0%; Score 9649; DB 9; Length 1863;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MDLSALRVEEVQVNVINAMOKITLPCICILELIKEPVSTKCDHIFCKFCMLKLLNOKKGPSQ	60
DB	1	MDLSALRVEEVQVNVINAMOKITLPCICILELIKEPVSTKCDHIFCKFCMLKLLNOKKGPSQ	60
QY	61	CPLCKNDITKRSLOESTRFSQVLELLKIIICAFQIDTGLEVANSYNPAKKENNSPEHLKD	120
DB	61	CPLCKNDITKRSLOESTRFSQVLELLKIIICAFQIDTGLEVANSYNPAKKENNSPEHLKD	120
QY	121	EVSIIISQGYNRNARKRLQSPENPSLOETSLVSQLSNLGTVRTLRTKQRTQPOKTSVYI	180
DB	121	EVSIIISQGYNRNARKRLQSPENPSLOETSLVSQLSNLGTVRTLRTKQRTQPOKTSVYI	180
QY	181	ELGSDSSEDTVNKATYCSVGQDQLLOITPOGTRDEISLDSAKKAACESEDTVNTTEHQ	240
DB	181	ELGSDSSEDTVNKATYCSVGQDQLLOITPOGTRDEISLDSAKKAACESEDTVNTTEHQ	240
QY	241	PSNNDLNTERRAARHPEKPYQSSVSNLHVPCGTNTHASSLQHSNSLLLTAKDRMNV	300
DB	241	PSNNDLNTERRAARHPEKPYQSSVSNLHVPCGTNTHASSLQHSNSLLLTAKDRMNV	300
QY	301	KAEPCKSKQGLARQSHNRWAGSKETCNDRRTPSTTEKKVDLNDADPLCERKENWKQLPC	360
DB	301	KAEPCKSKQGLARQSHNRWAGSKETCNDRRTPSTTEKKVDLNDADPLCERKENWKQLPC	360
QY	361	SNPRDTEDPWITLINSIQKVNWFERSDELGSDSDSHDGESESNKAKVADLDVLEVD	420
DB	361	SNPRDTEDPWITLINSIQKVNWFERSDELGSDSDSHDGESESNKAKVADLDVLEVD	420
QY	421	EYSGSSEKIDLLADPHEALICKSERVHKSVSNEIDKIFGKTYRKKASLPNLSHVYEN	480
DB	421	EYSGSSEKIDLLADPHEALICKSERVHKSVSNEIDKIFGKTYRKKASLPNLSHVYEN	480
QY	481	LIIGAFVTEPQIIQERPLTNLKKRRRPTSGLHPEDFIKKADLAVQKTPMINQGTNOTE	540
DB	481	LIIGAFVTEPQIIQERPLTNLKKRRRPTSGLHPEDFIKKADLAVQKTPMINQGTNOTE	540
QY	541	ONGQVNNITNSGHENKTKGDSIQNEKNPNPTESLEKESAPKTKAEPISSSISNMELELNI	600
DB	541	ONGQVNNITNSGHENKTKGDSIQNEKNPNPTESLEKESAPKTKAEPISSSISNMELELNI	600
QY	601	HNSKAPKNRLRRKSSSTRHIIHAELVVSRLNSPPNCTELQIDSCSSSEELKKKKYNOMP	660
DB	601	HNSKAPKNRLRRKSSSTRHIIHAELVVSRLNSPPNCTELQIDSCSSSEELKKKKYNOMP	660
QY	661	RHSRLQLMEGKEPATGAKSKNKNPEQTSKRHSDTPELKLTNAPGSGFTKCSNTSELKE	720
DB	661	RHSRLQLMEGKEPATGAKSKNKNPEQTSKRHSDTPELKLTNAPGSGFTKCSNTSELKE	720
QY	721	FVNPSPLEEKKEKLETVKVSNNAAEDPKDMLSGERVQLQTERSVESSSISLVPGTDYGTQ	780
DB	721	FVNPSPLEEKKEKLETVKVSNNAAEDPKDMLSGERVQLQTERSVESSSISLVPGTDYGTQ	780
QY	781	ESISLLEVSTLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDETEGFKYPLGHEVNH	840

DB	781	ESISLLEVSTLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDETEGFKYPLGHEVNH	840
QY	841	RETSIEMEESLDAQYLQNTFKVSKRQSFALFSPGNABEBCATFSAHSGSLKKQSPKVT	900
DB	841	RETSIEMEESLDAQYLQNTFKVSKRQSFALFSPGNABEBCATFSAHSGSLKKQSPKVT	900
QY	901	FECQKEENOGKNESNPKVOTVNTITAGFPVVGOKDKPVDNAKCSIKGGSRFCLSSQFRG	960
DB	901	FECQKEENOGKNESNPKVOTVNTITAGFPVVGOKDKPVDNAKCSIKGGSRFCLSSQFRG	960
QY	961	NETGLITPNKHGLLQNPYRIPPLFPKFSFKVTCKKNLLEENFEHSHMSPEREMGNENIP	1020
DB	961	NETGLITPNKHGLLQNPYRIPPLFPKFSFKVTCKKNLLEENFEHSHMSPEREMGNENIP	1020
QY	1021	STVSTISRNNTIRENVFKEASSNINEVGSSTNEVGSSINEIGSSDENTOAELGRNRGPKL	1080
DB	1021	STVSTISRNNTIRENVFKEASSNINEVGSSTNEVGSSINEIGSSDENTOAELGRNRGPKL	1080
QY	1081	NAMURLGVLOPEVYKQSLPGSNCKHPETKKOEYBEVQTVNTDFSPYLISDNLEQPMGSS	1140
DB	1081	NAMURLGVLOPEVYKQSLPGSNCKHPETKKOEYBEVQTVNTDFSPYLISDNLEQPMGSS	1140
QY	1141	HASQVCSETPDLLDDGGEIKEDTSAENDIKESSAVFSKQKGLSRSPSPFTHHLAQ	1200
DB	1141	HASQVCSETPDLLDDGGEIKEDTSAENDIKESSAVFSKQKGLSRSPSPFTHHLAQ	1200
QY	1201	GYRGAKKLESSEENLSEDEELPCFQHLHFGKVNINIPQSSTRHSTVATECLSKNTEENL	1260
DB	1201	GYRGAKKLESSEENLSEDEELPCFQHLHFGKVNINIPQSSTRHSTVATECLSKNTEENL	1260
QY	1261	LSLKSNDSCNQVILAKASQEHLSSETKCSASLFSQSCSELEDLTANTNQDFPLTGS	1320
DB	1261	LSLKSNDSCNQVILAKASQEHLSSETKCSASLFSQSCSELEDLTANTNQDFPLTGS	1320
QY	1321	SKQMRHOSQGVGLSDKELVSDDEERGTLGLENQBEQSDSNLGEAASGESETSVSE	1380
DB	1321	SKQMRHOSQGVGLSDKELVSDDEERGTLGLENQBEQSDSNLGEAASGESETSVSE	1380
QY	1381	DCSGLSSQSDILTTQORDTMOHNLIKLOQEMAEVLVEQHGQSPPNSYPSIISSSALE	1440
DB	1381	DCSGLSSQSDILTTQORDTMOHNLIKLOQEMAEVLVEQHGQSPPNSYPSIISSSALE	1440
QY	1441	DLRPEOSTSEKAVLTQKSEXPISQNPGLSADKPEVSADSSTSKNKEPVERSSPSK	1500
DB	1441	DLRPEOSTSEKAVLTQKSEXPISQNPGLSADKPEVSADSSTSKNKEPVERSSPSK	1500
QY	1501	CPSLDDRWYHSCSGSLQNRNYPQEEELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG	1560
DB	1501	CPSLDDRWYHSCSGSLQNRNYPQEEELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG	1560
QY	1561	TPYLESGISLFSDDPESDSDRAPESARVGNIPSSTSALKVPQLKVAESAQSAHAHTT	1620
DB	1561	TPYLESGISLFSDDPESDSDRAPESARVGNIPSSTSALKVPQLKVAESAQSAHAHTT	1620
QY	1621	DTAGYNAMESVSREKPELTASTERVNKRMSVSGLTPEEFMLVYKFAKHHTLTNLI	1680
DB	1621	DTAGYNAMESVSREKPELTASTERVNKRMSVSGLTPEEFMLVYKFAKHHTLTNLI	1680
QY	1681	TEETHVVMKTDADFVCGERTLKYFLGAGGKVVVSYFWVTQSIKERKMLNHDHFEVRGDV	1740
DB	1681	TEETHVVMKTDADFVCGERTLKYFLGAGGKVVVSYFWVTQSIKERKMLNHDHFEVRGDV	1740
QY	1741	VNGRNOGPKRARESQRKIFRGLIEICCYGPFTHMPTDQLEWMVOLCGASVVKELSSFTL	1800
DB	1741	VNGRNOGPKRARESQRKIFRGLIEICCYGPFTHMPTDQLEWMVOLCGASVVKELSSFTL	1800
QY	1801	GTGVHPITVVQPDADWEDNGFHAIGQCEAPVTVREWVLDVSVALYQCOELDTYLPQIPH	1860
DB	1801	GTGVHPITVVQPDADWEDNGFHAIGQCEAPVTVREWVLDVSVALYQCOELDTYLPQIPH	1860
QY	1861	SHY 1863	

|||

Db 1861 SHY 1863

RESULT 2

US-09-734-672-2
; Sequence 2, Application US/09734672
; Publication No. US20020183268A1
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; Allen, Antonette C.
; Alvares, Christopher P.
; Critz, Brenda S.
; Olson, Sheri J.
; Schelter, Denise B.
; Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; BRCA1 Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan Lewis & Bockius LLP
; STREET: 1111 Pennsylvania Ave., N.W.
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/734,672
; FILING DATE: 03-Dec-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/966,436
; FILING DATE: 07-No. US20020183268A1-97
; APPLICATION NUMBER: US 08/598,591
; FILING DATE: 12-Feb-96
; ATTORNEY/AGENT INFORMATION:
; NAME: Michael S. Tuscan
; REGISTRATION NUMBER: 43,210
; REFERENCE/DOCKET NUMBER: 44921-5055-02-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-739-3000
; TELEFAX: 202-739-3001
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: No. US20020183268A1 Relevant
; TOPOLOGY: No. US20020183268A1 Relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-734-672--2

Query Match 99.9%; Score 9635; DB 9; Length 1863;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1860; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MDLSALRVEEVQNVINAMQKILECPICLEIKPEVSTKCDHIFCKFCMLKLLNQKGPSQ 60
Db 1 MDLSALRVEEVQNVINAMQKILECPICLEIKPEVSTKCDHIFCKFCMLKLLNQKGPSQ 60
Qy 61 CPLCKNDITKRSLOESTRFSQVLEELKIIICAFQDITGLGYANSYNPAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVLEELKIIICAFQDITGLGYANSYNPAKKENNSPEHLKD 120

Qy 121 EVSIIQSMGYRNRAKRLQLQSEPNPSLQETSLSVQLSNLGTVRTLRTKQIQPKTSVYI 180
Db 121 EVSIIQSMGYRNRAKRLQLQSEPNPSLQETSLSVQLSNLGTVRTLRTKQIQPKTSVYI 180
Qy 181 ELGSDSSDVTNKTATYCSVGDQELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSDVTNKTATYCSVGDQELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Qy 241 PSNDLNTTEKRAAERHPEKYQGVSSVSNLHVEPCGTTNTHASSLQHENSLLLTKDRMNVE 300
Db 241 PSNDLNTTEKRAAERHPEKYQGVSSVSNLHVEPCGTTNTHASSLQHENSLLLTKDRMNVE 300
Qy 301 KAFCNKSKQPLARSQHNWAGSKETCNDRRTPSTPEKKYVDLNADPLCERKWNKQLPC 360
Db 301 KAFCNKSKQPLARSQHNWAGSKETCNDRRTPSTPEKKYVDLNADPLCERKWNKQLPC 360
Qy 361 SENPRDTEVPWITLSSIOKVNEWFSRDELGLGSDSHDGESESNKAVADVLVLENYD 420
Db 361 SENPRDTEVPWITLSSIOKVNEWFSRDELGLGSDSHDGESESNKAVADVLVLENYD 420
Qy 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVSNEIEDKIFGKTYRKASLPNLSHVTE 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVSNEIEDKIFGKTYRKASLPNLSHVTE 480
Qy 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINQGTNOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINQGTNOTE 540
Qy 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNWELENI 600
Db 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNWELENI 600
Qy 601 HNSKAPKKNLRKRSRTHALELVVSRNLSPNCTELOIDSCSSSEELKKKYNOMPV 660
Db 601 HNSKAPKKNLRKRSRTHALELVVSRNLSPNCTELOIDSCSSSEELKKKYNOMPV 660
Qy 661 RHSRNLQMEGKEPATGAKKSNKPNEQTSKRHSDTPELKLTNAPGSFTKCSNTSELKE 720
Db 661 RHSRNLQMEGKEPATGAKKSNKPNEQTSKRHSDTPELKLTNAPGSFTKCSNTSELKE 720
Qy 721 FVNPSPREEKEKLETVKYSNNAEDPKDMLSGERVLOQTERSVESSISLVPGTDYGTQ 780
Db 721 FVNPSPREEKEKLETVKYSNNAEDPKDMLSGERVLOQTERSVESSISLVPGTDYGTQ 780
Qy 781 ESTSLLEVSTLGRAKTEPNKCVSOCAAFENPKGLIHGCCSDNRNDTEGFKYPLGHEVNH 840
Db 781 ESTSLLEVSTLGRAKTEPNKCVSOCAAFENPKGLIHGCCSDNRNDTEGFKYPLGHEVNH 840
Qy 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFNSPGNAEECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFNSPGNAEECATFSAHSGSLKKQSPKVT 900
Qy 901 FECEQKEENOGKNESNIKPVQTNITAGFPVVGOKDKPVDNAKCSIKGSRFCFLSSQFRG 960
Db 901 FECEQKEENOGKNESNIKPVQTNITAGFPVVGOKDKPVDNAKCSIKGSRFCFLSSQFRG 960
Qy 961 NETGLIIPNKHGLLQNPYRIPPLFPPIKSFVKTKCKNLLLENFEHSHMSPEREMGNINIP 1020
Db 961 NETGLIIPNKHGLLQNPYRIPPLFPPIKSFVKTKCKNLLLENFEHSHMSPEREMGNINIP 1020
Qy 1021 STVSTISRNINIRENVFKAEASSNINEVGSSTNEVGSSTNEIGSSDENIQALGNRGPKL 1080
Db 1021 STVSTISRNINIRENVFKAEASSNINEVGSSTNEVGSSTNEIGSSDENIQALGNRGPKL 1080
Qy 1081 NAMLRGLVQPEVYKQSLPGSNCKHPEIKKOEYEEVQTVNTDFSPYLIISDNLEQPMGSS 1140
Db 1081 NAMLRGLVQPEVYKQSLPGSNCKHPEIKKOEYEEVQTVNTDFSPYLIISDNLEQPMGSS 1140
Qy 1141 HASQVCSETPDDLLDDGEIKEEDTSAENDIKESSAVFSKVOKGELSRSPPTHTTHAQ 1200
Db 1141 HASQVCSETPDDLLDDGEIKEEDTSAENDIKESSAVFSKVOKGELSRSPPTHTTHAQ 1200
Qy 1201 GYRRGAKKLESSEENLSSEDEELPCFQHLLFGKVNINIPSPQSTRHSTVATECLSKNTEENL 1260

Db 1201 GYRRGAKLSESENLSSDEELPCFQHLFLCKVNNIPQSQRHSTVATECLSKNTEENL 1260
QY 1261 LSLKSLNDCSNQVILAKASQEHLLSEETKCSASFSSQCSLEDLTANTNQDPFLTGS 1320
Db 1261 LSLKSLNDCSNQVILAKASQEHLLSEETKCSASFSSQCSLEDLTANTNQDPFLTGS 1320
QY 1321 SKOMRHOSESQGVGLSDKELVDDEERGTLLENNOEQQSDNLEGEASGCESETSYSE 1380
Db 1321 SKOMRHOSESQGVGLSDKELVDDEERGTLLENNOEQQSDNLEGEASGCESETSYSE 1380
QY 1381 DCSGLSSQSDILTTQQRDTMQHNLKLOQEMAELEAVLEQHQSQSPNSYPSTIISDSSALE 1440
Db 1381 DCSGLSSQSDILTTQQRDTMQHNLKLOQEMAELEAVLEQHQSQSPNSYPSTIISDSSALE 1440
QY 1441 DLNRPQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPVERSSPSK 1500
Db 1441 DLNRPQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPVERSSPSK 1500
QY 1501 CPSLDDRWYMHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDDRWYMHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESGISLFDPSDPSSEDRAPEARVGNIPSTTSALKVPQLKVAESAGQPAAAHTT 1620
Db 1561 TPYLESGISLFDPSDPSSEDRAPEARVGNIPSTTSALKVPQLKVAESAGQPAAAHTT 1620
QY 1621 DTAGYNAMESVSRKEPELTASTERVNRKMSMVVSGLTPEEFMLVYKFAKHHITLTNLI 1680
Db 1621 DTAGYNAMESVSRKEPELTASTERVNRKMSMVVSGLTPEEFMLVYKFAKHHITLTNLI 1680
QY 1681 TETTHVVMKTDAEFCERTLAYFIAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
Db 1681 TETTHVVMKTDAEFCERTLAYFIAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRHHQGPRAESQDRKIFRGLGICCGYPPTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
Db 1741 VNGRHHQGPRAESQDRKIFRGLGICCGYPPTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
QY 1801 GTGVHPVVVQPDANTEDNGFHAIGQMEAPVVTREWVLDVALYQCOELDTYLIPQIPH 1860
Db 1801 GTGVHPVVVQPDANTEDNGFHAIGQMEAPVVTREWVLDVALYQCOELDTYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 3

US-09-734-672-6
; Sequence 6, Application US/09734672
; Publication No. US20020183268A1

GENERAL INFORMATION:

APPLICANT: Murphy, Patricia D.
Allen, Antonette C.
Alvares, Christopher P.
Critz, Brenda S.
Olson, Sheri J.
Schelter, Denise B.
Zeng, Bin

TITLE OF INVENTION: Coding Sequences of the Human

BCRAL Gene

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESS:

ADDRESSEE: Morgan Lewis & Bockius LLP
STREET: 1111 Pennsylvania Ave., N.W.
City: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/734,672
FILING DATE: 03-Dec-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/966,436
FILING DATE: 07-No. US20020183268A1-97
APPLICATION NUMBER: US 08/598,591
FILING DATE: 12-Feb-96
ATTORNEY/AGENT INFORMATION:
NAME: Michael S. Tuscan
REGISTRATION NUMBER: 43,210
REFERENCE/DOCKET NUMBER: 44921-5055-02-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-739-3000
TELEFAX: 202-739-3001
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: No. US20020183268A1 Relevant
TOPOLOGY: No. US20020183268A1 Relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BCRAL
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-734-672-6

Query Match 99.9%; Score 9635; DB 9; Length 1863;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1860; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQNVINAMOKILECPICLELKEPVSTKCDHIFCFCKMLKLLNOKKGPSS 60
Db 1 MDLSALRVEEVQNVINAMOKILECPICLELKEPVSTKCDHIFCFCKMLKLLNOKKGPSS 60
QY 61 CPLCKNDITKRSLOESTRFSQVVEELLKIIICAFQDITGLGYANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVVEELLKIIICAFQDITGLGYANSYNFAKKENNSPEHLKD 120
QY 121 EYSIIQSMGYNRNRAKRLQSEPNPESLQETSLSVQSLNLTGVTTLTKORIQPKTSVYI 180
Db 121 EYSIIQSMGYNRNRAKRLQSEPNPESLQETSLSVQSLNLTGVTTLTKORIQPKTSVYI 180
QY 181 ELGSDSSEDVTNKATYCSVGQDQLQITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDVTNKATYCSVGQDQLQITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPEKYQGSVSNLHVPCGTNTNTHASSLQHENSSLLLTKDRMNYE 300
Db 241 PSNNDLNTTEKRAAERHPEKYQGSVSNLHVPCGTNTNTHASSLQHENSSLLLTKDRMNYE 300
QY 301 KAFCNKSQKPLARSQHNWAGSKETCNDRTPTSTKRVLDNADPLCERKEWNKQKLPC 360
Db 301 KAFCNKSQKPLARSQHNWAGSKETCNDRTPTSTKRVLDNADPLCERKEWNKQKLPC 360
QY 361 SENPRDTEVPWITLNSIQKWNWFSRDELLGSDSDHGESESNKAVADVLDVLYNEVD 420
Db 361 SENPRDTEVPWITLNSIQKWNWFSRDELLGSDSDHGESESNKAVADVLDVLYNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHVTEEN 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHVTEEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPDEFKTKADLAVQKTPPEMNOGTNOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPDEFKTKADLAVQKTPPEMNOGTNOTE 540

; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12996
 ; LENGTH: 6281
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-12996

Query Match 3.5%; Score 341; DB 10; Length 6281;
 Best Local Similarity 18.3%; Pred. No. 1.1e-09;
 Matches 335; Conservative 317; Mismatches 776; Indels 404; Gaps 79;

QY 7 RVEEONVINAMOKILEPICLELIKREPVSTKCDHIFCKMLKLLNKKGPSCQCLCKN 66
 Db 2848 QVEAAMNOVNATAALNGTONLEKAKQHANTAID-----GLSHLTNAQK 2891
 QY 67 DITKRSLOESTRESQVLEELKIIICAFOLDTGLE-----YANSYNFAKKNNSP 115
 Db 2892 EALKQLVQOSTTVAEAGGNEQ--ANNVDAAMDKLRQSIADNATTKQNGNYTDASQNK 2948
 QY 116 EHLKDEVSIIQSGYRNRKRLLQSPENPSLOETSLVSOLNGLTVR-TLRTKQRIQ 174
 Db 2949 DAYNNAVTTAAG-----IIDQTSPTLPTVINQAAGQVSTTKNALNGNENLEA 2998
 QY 175 KTSVYELGSDSSEDTVNKATYCSVGQ-----ELQITPQGTREDEISLDSAKKAC 226
 Db 2999 KQASOSLG---SLDLNNAKQOTVTDQINGAHTVDEANQIKQANLNTAMGNLQOIA 3055
 QY 227 EFSETDVTNTEHHOPSNNDLNTTEKRAAERHPEKYOGSSVSNLHVPPCGTNTNTHASSLQHE 286
 Db 3056 DKDATRAT-----VNETDAQKQOAYNTAVTNAENIISKANGGNATQAEVEQA 3104
 QY 287 NSLLLTIDKRMNVEKAEFCNKKQOPGLARSOHNRWAGSKETCNDRTTPSTEKKVDLNADP 346
 Db 3105 IKOVNAKQALN-----GNANVQH-----AKDEATALINSNDLNQ-- 3140
 QY 347 LCBKRWKOKLPCSENPRTDVPWITLNSSTQKWNEMFSDSDELLGSDSDHGESEN 406
 Db 3141 --AQKALKQOVQNAVTVAGVNV-----KQTAQELNNAMTQLKQ--GIADKEQTKADGN 3191
 QY 407 AKVADVLVNLVDEYSGSEKIDLLASDPHEALICKSERVSHKSVESNIEDKIFGKTYR 466
 Db 3192 FVNADP-----DKQNAVQAVAKAEALISATPDVVVTPSEITAAALNKVYQAKNDLNGNTNL 3247
 QY 467 KKA-----SLPNLSHV-----TENLIIGAFVTEPQIIQERPLT-----NKLKR--- 504
 Db 3248 ATAKQNVQHAIDQLPNLNAQRDEYSKQITQATLVPVNAIQQAATTLNDAMTQLKOGIA 3307
 QY 505 ---KRRPTSLGHPEDFIKKA--DLAVQKTPEMINOCTNOT-----EONGQVMNITNSG- 552
 Db 3308 NKAQIKGSYENYHADTDKQATYDANATYKAEELKQNTNPTMDPNTIQOALTQVNDTNQAL 3367
 QY 553 HENKTKGDSIQNEKNP-NPIESLE--KESAFKTKAE--PISSSISNMELEL-NIHNKAP 606
 Db 3368 NGNOKLADAKQADAKTTGLTLDHLNDAKQALTTQVEQAPDIATVNVNKKQNAQLNNAMTN 3427
 QY 607 KKNRLRKRSTRIHALELVVSRNLSPNCPNCTELOIDSCSSSEBIKKKKYNOMPVRHSRL 666
 Db 3428 LNNALQDKTET-----LNSINFTD-----ADQAKKDAYTN-AVSHAEGI 3465
 QY 667 QLMGKEPATGAKKSNPEQTSKR-----HSDTTPPELK-----LTNAPGSGFTKC 712
 Db 3466 -----LSKANGSNASQTEVEQAMQVRNEAKQALNGNDNVQRAKDAKQVITNA----- 3513
 QY 713 SNTSELKEFVNPSPREEKEKLETY----KVSNNAEADPKDMLSGERVLTQTSRVSSS 768
 Db 3514 ---NDLNQAKQDAL--KQOVDAQTVANVTIKQTAQDLNQAMTQLKQGIADKQDKANG 3568
 QY 769 ISLVPGTD----YGTQESISLLEVSTLIGKAKTEPKNCVSOCAAFENPKGLIHG-----CS 819
 Db 3569 NFYNADTDKQANVNAVAHAEQIISGTNPANVDPQOVAQALQOVNAQKGLDNGNHNLQVA 3628

QY 820 KDNRNDETEGFKYPLGHEVNHVSRETSTEMEESSELDQYQLNTFKVSKRQSFALFSPNGAE 879
 Db 3629 KDNAN-TAIDQLP--NLNQPKTALKQDVSHAE---LVTGVNAIKQNDALNANMGTUK 3681
 QY 880 ECATFSAHSGSLKKSPKVFCECEKEENQOGKSNIRKPVQTVNITAGFP-----VV 932
 Db 3682 QQ-----IQANSQVQSVDFT-QADQDKQOAYNNAAN---QAAQIANGIPTVPLTPDVT 3731
 QY 933 GQKDKPYDNAKCSIKGSRFCJSSQ--FRGNET--GLITPNKHGLLQNPYRPPLEPIKS 988
 Db 3732 TQAVTTMNAQKDALNGDEKLAQAKQALANLDTLRLNQPQDRLRNQINQAALATVEQ 3791
 QY 989 FVTKTKCKNLEENFEHSMSPERMGNE--IPSTVSTISNNI--RENFKKASSSN 1043
 Db 3792 TKQNAQNVNTAMSNLKGATANKDTVKASENYHDADADKQATYNAVSAQEGIINTNTPT 3851
 QY 1044 IN--EYGSSTNEVGSSINEIGSSDENIQAPLGRNRPKLNAMLRLGLVLOPEVYKQSLPGS 1101
 Db 3852 LNPDEITRALTOVTDKLN--GLNGEAKLATEKQNAKDAVSGMTHLN---DAQKQALKQ 3905
 QY 1102 NCKHPEIKQVEYEVQTVNTDPSYLIISDNLEQPMGSSHASQVCSSETPDDLLDGEIKE 1161
 Db 3906 IDQSPET--ATVNOVKQT-----ATSLDQAM--DQLSQAINDKAQTLADGNYLNA 3951
 QY 1162 DTSFAENDIKE-----SSAVESKSVQKGLSRSPSPFTHLQAGYRRGAKKLESSSENLS 1217
 Db 3952 DPD-KONATYKQAVAKAEALLNKQSGTNEVQAVESTINEVNA-----AKQALNGNDNIA 4004
 QY 1218 S-----EDEELPCFQHLHFGKVNINPQSQSTRHSTVATECLSKNTEENLLS-----LKNSLN 1268
 Db 4005 NAKQAKQQLANLTHLNDAAKQSFESQITQ-APLVTDVITINQKQATLDHAMELLRNSVA 4063
 QY 1269 DCSNVILAKASQEHLSSETKCSASLFSQSCSELEDLTANTNTQDPFLLIGSKQMRHOS 1328
 Db 4064 D--NOTTL--ASDYH-----DATAQRQNDYNOAVTAANNINOT 4099
 QY 1329 ESQGVGLSD-----KELVSDDEBRTGTLENNQOEQSMDSNLGEAASGCESETS 1377
 Db 4100 TSPMNPDDVNGATTOVNTKVALDGDENLAAKQKQANNRLDOLDHLNNAQKQQLQSQIT 4159
 QY 1378 VSEDCGLSSQSDILTQQRDTMQLNHLIKLOQEMAELEAVLEQHG-----SQPSNSYPS 1431
 Db 4160 QSSDIAAVNGHKQ--TAESLNTAMGNLI---NAIADHQAV-EQGRGNFINADTKQFAYNT 4213
 QY 1432 IISDSAL--EDLRNPEQSTSEKAVLTQKSSSEYIPSONPEGLSAD-KFEVSADSSTSK 1487
 Db 4214 AVNEAAMINKQTQONANQTEVEQAITYQTTL-----QALNGDHLQVAKTNATQ- 4264
 QY 1488 NKEPGVGRSPSKPCLDDRWYMHSCSGSLONRNYPQSELIKVVDEE-----QOLEES 1542
 Db 4265 -----AID-----ALTSLNDQKQALKDQVTAATVLTAVHOIQN 4299
 QY 1543 GPHDLTETSLPRQDLLEGTPLYESGLSFLSDPESDPSEDRAPEASRVGNIPSSTSALKV 1602
 Db 4300 A-NTLNAQMHGLRQSIQDNNATKANSKYINEDQPEQONDQAVQAA--NNIINEQTA--- 4353
 QY 1603 POLKVAESAQSPAAAHITDTOTAGYNAMEEVSREKPELTASTERVNKR--MSWVVSGLTPE 1660
 Db 4354 -----TLDNNAINAATVNTTKAALHGDVQLQNDKHAKQTVSOLAH- 4396
 QY 1661 EFMVYKFAKKHHITLTNLITETTHVMKTD 1692
 Db 4397 -----LNNQAKH--MEDTLIDSETTFTAVKQD 4421

RESULT 6
 US-09-815-242-5635
 ; Sequence 5635, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.

```

; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELTRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5635
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5635

```

```

Query Match          3.48; Score 328.5; DB 10; Length 2368;
Best Local Similarity 18.28; Pred. No. 1.3e-09;
Matches 349; Conservative 306; Mismatches 720; Indels 547; Gaps 78;

```

```

QY 93 FOLDGTGLVANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRA-----KRLQSE 141
DB 390 FINT--EIGNNGFG--ASLKAQDFKEVTLPGQVYVNNLSLTTFPNGNEDSTVLKNN 445
QY 142 PNPISQTSLSVOLNSLGTVRTLTQRIOPOKT---SVYELGSDSSEDTVN---KAT 195
DB 446 TVNYDQNAKVTFTSQGVTYARGTHTKEVLPDPKSLSKYKVNANIDTPKIDFNEKLT 505
QY 196 YCSVGDQELLOITPOCTRDELSLSAK--KAACFSETDVTNTEHHQPSNNDLNTTKRAA 254
DB 506 YRTASD-----IVINNAQPEVLTADPFSVAVEMNKDALQQQVNSQVDNSHYTTAS--IA 558
QY 255 ERHPEKYQGSVSN---LHVPEPC-----GTNTHASSLQHENSSLLLTKDRMVEKA 302
DB 559 EYNKLQQAADNLNEDANHVETANRASAQAAIDGLVTKLQALIDNQAALAEIDAKAQEKV 618
QY 303 EFCNKSQPGFLARSOHNRWAGSKTCNDRRTPPTEKKVDLNADPLCERKE-----WNKQK 357
DB 619 TAAQGSKVV---TQDEVAALVTIINNKKNAIAEINKQTAAQGVTTKONGIAVLDDQV 674
QY 358 LPCSENPRDTEVPWITLNSSIQKWNWFSSRDELLSGSDSHDGESESNKAVADVLDVLN 417
DB 675 IPTPVKPOAKODI-----IQAVT---TRKQOI-----KKSNASLQDEKDVAN 713
QY 418 EYDEYSGSSEKIDLLASPHPELACKSERVHSKSVESNIEDKIFGKYRKKASLPNLSHV 477
DB 714 D-----KIGIKETKAIDDAATNQAQVEAKTKALINDINQATAPATAKAAALEEDEV 767
QY 478 TENLI-----IGAFVTEPQIIQERPLTNKLKRKRPTSLGHPEDFIKKADLAVOKTPMIN 533
DB 768 VQAQIQDAPLNPDTTNEEVAEIERINAAK-----VSGVK-----AIEATTTAQD 812
QY 534 OGTNOTEONGVMNTNSGHENKTKGDSIQNEKNPNFIESLEKESAFKTKAEPISSISIN 593
DB 813 LERVKNEEISKIENTDS---TOTKMDAYNEVK-----QAATARKTQATVNSNATNE 861

```

```

QY 594 MELELNHNKAPKXKNRLRRKSSSTRHHALELVVSR-----NLS 632
DB 862 EVAEADAARAAQKQG-----LHDIQVVKSKQEVADTKSKVLDKINAQTQAKVK 911
QY 633 PPNCT-----ELQIDSCSSSEIEIKKKYKQMPVHRSR---NLQLMGKGPATGA 678
DB 912 PAADTEVENAYNTRKQETQNSNASTTEE-KQAAYTELDTKKQEARTNLDANTNSDVTTA 970
QY 679 KKS-----NKPNEOTSKRHSDTPELKLITNAPGSFTKCSNTSELKRVNPSLPREEK-- 731
DB 971 KNGIAAINQVQAATTK--SDAKAEIA-----QASERKTAIEAMNDSTTEEQAA 1020
QY 732 EEKLETYKVSNNADPKDMLSGSERVLQTERSVESSISLVPDGTGYGTOESISLLE---- 787
DB 1021 KDKVDQAVVTANAD--IDNAAANTVDNNAKTITNEATITPDANVKPTAKQAIADKVA 1078
QY 788 ----VSTLGRKAKTEPNKCVSQCAAFENPKGLIHGCSKNRNDTEGFKYPLGHEVNHRET 843
DB 1079 QETAIDANNGATTEEKAQKQVQTEK-----TTADTAID-----GAHTNAEVEA 1123
QY 844 SIEMEESELDAOYLQNTFKVSKROSEALFSPG-----NAEBECATFSAHSGSLK 893
DB 1124 AKNAEIAKIEAIQATTTKDNAKQAIATKANERKTAITAOQDITAEETAAANAVDNVAT 1183
QY 894 KQSPKVTFECEQKEENQK--NESNIKPV-OTV-----NITAGFPVVGQKDKP 938
DB 1184 QANNIEAANSQNDVDQAKTTGEASIDQVTPVNKKATAVTDAKNNITA-----ATDDNG 1238
QY 939 VDNK-----CSIKGGR-----FCLSSQFRNETGLITPNKHL----- 973
DB 1239 VDTAKDAGKNSIQSTQPATAVKSNKNDVDQAVTTQQAIDNTTGTATTEERNAARDVLK 1298
QY 974 -----LQNPYRIPPLPRIK-----SFVKTCKKNLLENEEHSM---- 1008
DB 1299 AKEKAYODILNAQTNDVTQIKQOAVADVQGITADTTIKQVAKDELATKAREQKALIAQT 1358
QY 1009 -----SPERE-----MGNEIPSTVSTISRNIRENVFKEASSNINEVGSSTN- 1052
DB 1359 ADATTEKEQANOQVDAELTOGQNIENAOISDIDVNTAKDNAIOA-----IDPIOASTDV 1413
QY 1053 -----EVGSSINEIGSSDENIOAELGRNRP-----KUNAMRLG-----V 1088
DB 1414 KTNARABELTEMQNKITEILNNNETTNEEKGNDIGPRAAYEEGLNNINAATTTGDVTTA 1473
QY 1089 LQPEVYK-QSLPGSNCKHPEIK-----KOEYEVVQTVNTDPS 1126
DB 1474 KDTAVQVQOQLHNPVKRPKAGKTALDQAAADKTKQIEOTPNASQOEINDAKQEVDTLQ 1533
QY 1127 YLISDNLEQPMGSSHASQVCSPTDDLLDDGE-----IKEDTSAENDIKESSAVFSKV 1181
DB 1534 --AKTNIDQ-----SSTDEYVDNAVKEGKAKINAVKTFSEYKKDALAKIEAAYNAKV 1583
QY 1182 QKGLSRSPSPFTHLAAQGYRGAKKLESEENL-----SSEDEELPCFQHL----- 1229
DB 1584 TEADNSNAS---TSSEIAEAKQKLAELKQATADQNVNQAATSKDDIEVQIHNDLNDYTI 1640
QY 1230 -----LFG-----KVNIPQSQR---HSTVATECLSKNTEENLLSLKSLND- 1269
DB 1641 PTGKESATDLYAYADQKKNISADTNATQDEKQQAQKQVDQNVQTALESINNGVDNGD 1700
QY 1270 -----CSNQVILAKASQEHLSSETKCSASLFSQCSSEDL 1306
DB 1701 VDDALTOGKAIDAIOVDATVKPKANQVIDAKA-----EETK-----ESIDQSDQL 1746
QY 1307 TANTQDQPELIGSSKQMRHOSQGVGLSKDELVSDEE--RGTGLE-----ENNOQE 1358
DB 1747 TAEKTEALAMI---KQITDQAKO---GITDATTTAEVERAKAQGLEAFQNDIQIDSTEKQ 1800
QY 1359 QSMDSNLGEARASCESETSVSEDCSGLSSQSDILTTQORDTMQHNL-----IKLQ 1408
DB 1801 KALE-----ELETALDQIEAGVNVQAD-ATTEKEAFTNALEDILSKATEDISQ 1849
QY 1409 QEMAELEAV----LEQHGQSQPSNSYPSIISDSSALEDLRNPESQSTSEKAVLTOKSESEYP 1464

```

Db 1850 TTNAETATVKSALBQLAQRINP-----VVKNALEAIREVNVKQIEIINKADADASAKE 1905
Qy 1465 ISONPEGLSADKFEVSADSSKNKE-----PGVERSSPSKCPSLDDRWYHMSCG 1515
Db 1906 IARTDLGRYDFRFBADKLD-KTQNTIEVABLQNTWTPAIEAIVPQNDPNAND-----TNSG 1959
Qy 1516 S-----LQNRNPSQBELIKVVDV----- 1534
Db 1960 SONNDATANSANATPENTGQPNVTESTDNANADTSSTTTNNQNDAACTTATSANSSA 2019
Qy 1535 -----EEQOLESSEGHDTETSYLPRQDLEGPYLESIGLSISLSDPESDPSEDPAPEASR 1589
Db 2020 TTDANDKFOANNSSADTSTNSPTMDNDVTSKPEVESTNNGTTDKPATEADNATPAESAT 2079
Qy 1590 VGNIPSSTSALKVPOLKVAESAQSPAAAHHTTDTACYNAMESV-----SREKPELTAS 1642
Db 2080 NNNSTTTATNENAP-----TESTATPTTASTGAESSADSKDNASVNDSKONAEVNS 2132
Qy 1643 TE 1644
Db 2133 AE 2134

RESULT 7

US-09-815-242-12389
; Sequence 12389, Application us/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12389
; LENGTH: 2368
; TYPE: prt
; ORGANISM: Staphylococcus aureus
US-09-815-242-12389

Query Match 3.4%; Score 328.5; DB 10; Length 2368;
Best Local Similarity 18.2%; Pred. No. 1.3e-09;
Matches 349; Conservative 306; Mismatches 720; Indels 547; Gaps 78;

Qy 93 FOLDNGLEVANSYFNAKKENNSPEHLKDEVSIIQSMGYRNA-----KRLQSE 141
Db 390 FTINT--EIGNNGNFG--ASLKADFKYEVLTPQGVTVVNSLTTFFPNGNEDSTVLKNN 445
Qy 142 PENPSLQETSLSVQLSNLGTVRTLTQRIOQPKT---SVVIELGSDSSEDVTN---RAT 195

Db 446 TVNYDONANKVFTTSGQVTTARGTHTKEVLPDPKSLKSYKVNANIDTPKNIDENEXTL 505
Qy 196 YCSVGDQELLQITPOGTDRDEISLDSAK-KAACEFSETDVTNTEHHQPSNNDLNTTEKRAA 254
Db 506 YRTASD-----IVINNAQPEVTLTADPFSVAVEMKMDALQQOVNSQVDSNHYTTAS--IA 558
Qy 255 ERHPKPYGSSVSN---LHVEPC-----GTNTHASSLQHENSLLLTKDRMNVEKA 302
Db 559 EYNKLIKQADNILLNEDANHVEATANRASQAADGLVTKLQAAALIDNQAIAELDAKAQEKV 618
Qy 303 EFCNKSKQGLARSOHNRWAGSKETCNDRRTPSTPEKKYVDLADPLCERKE-----WNKOK 357
Db 619 TAAQOSKKV---TQDEVAALVTKINNDKNAIAEINKQTTAQGVVTEKDNGLAVLDQDV 674
Qy 358 LPCSENPROTEDVPMTILNSSITQKYNWFSDLLSGDSDHSDGSESNKAVADVLDVLN 417
Db 675 TPTVKPQAKODI-----IQAVT---TRKQQT-----KKSNASLQDEKDVAN 713
Qy 418 EVDEYSGSSEKIDLLADPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHV 477
Db 714 D-----KIGKIETKAIKDIDAATTNAQVEAIKTAINDINQOTAPATTAKAAALEEDEV 767
Qy 478 TENLI-----IGAFVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVQKTPPEMIN 533
Db 768 VQAQIDQAPLNPDDTTNEEVAEAIERINAAK-----VSGVK-----AIEATTTAQD 812
Qy 534 QGTNQTQEQGVMMNITNSGHENKTKGDSIQNEKNPNPESLEKESAFKTKAEPISSSISN 593
Db 813 LERVKNEEISKIENITDS---TQTKMDAYNEVK-----QAATARKTQNTATVSNATNE 861
Qy 594 MELELNHNHNSKAPKKNRLRRKSSSTRHIALELVSR-----NLS 632
Db 862 EVAEADAVERAAQKOG-----LHDIVVKSQEVADTKSVKLDKINAIQTQAKVK 911
Qy 633 PPNCT-----ELQIDSCSSSEIEKKKKYNQMPVHRH---NQLMEGKEPATGA 678
Db 912 PAADTEVENAYNTRKQEIQNSNASTTEB-KQAAAYELDTKKQEARTNLDAANTNSDVITA 970
Qy 679 KKS-----NKPNQTSKRHSDSTFPPELKLITNAPGFTKCSNTSELKEFVNPSLPREEK-- 731
Db 971 KNGIAAINOVQAAATTK--SDAKAEIA-----QKASERKTAIEAMNDSTTEEOQAA 1020
Qy 732 EEKLETVKVSNNAEDPKDMLSGERVLTQERSVESSSISLVPDGTDTGTQESTISLE---- 787
Db 1021 KDKVDQAVVTANAD--IDNAAANTVDNAKTTNEATIAITPDANVKPTAKQAIAADKVOA 1078
Qy 788 ----VSTLGKAKTEPNKCVSQCAAFENPKGLHGCSDKNRNDTEGFKYPLGHEVNHRSRET 843
Db 1079 QETAIDANNGATTEEKAQKQVQTEK-----TTADTAID-----GAHTTNAEVEA 1123
Qy 844 SIEMEESELDQYLONTFKVSKRQSFALFNPQ-----NAAEECATFSAHSGSLK 893
Db 1124 AKNAEIAKIEAIOPATTTTKDNAKQAIATKANERKTAIAQTODITAEETAAANANVDNAT 1183
Qy 894 KOSPKVTECECKEENOCK--NESNIKPV-OTV-----NITAGFPVVCQKOKP 938
Db 1184 QANNIEAANSQNDVDQAKTTGEASIDQVTPTVNKKATAVTDAKNNITA-----ATDDNG 1238
Qy 939 VDNAK-----CSIKGGSR-----FCLSSQFRGNETGLITPNKHGL----- 973
Db 1239 VDTAKDAGKNSIQSTQPATAVKSNNAKNDVDQAVTTQNCQIDNTTIGATTTEKNAAKDLVLK 1298
Qy 974 -----LQNPYRIPPLFPPIK-----SFVKTKCKKNLLEENFEESHM----- 1008
Db 1299 AKEKAYQDILNAQTNDVTDQKQAVADVQGITADTTIKDKVAKDELATKAREQKALIAQT 1358
Qy 1009 ---SPERE-----MCNENIPSTVSTISNNRENVPKREASSNSNEVSSNTN- 1052
Db 1359 ADATTEEQANQAVDAELTQCNQNIENAQSIDDVNTAKDNAIQ-----IDPQASTDV 1413
Qy 1053 -----EVGSINSIEIGSDENIQAEELGNRNP-----KLNAMLRIG-----V 1088

Db 697 RISSRKK-----KDPSPKVLPEPNKQBKTEKEETNVR 732
QY 922 TVNITAGPVVQGDKPDVNAKSIKGSRFCLSSQFRGNETGLITPNKKGHLQNPYRIP 981
Db 733 TLRSRPRISRTAKVAEIROQADKKKGEG---EDEVESSTALQTKDKKEIL-----782
QY 982 PLFPKISVTKCKKNLEENFEHSMSP-----REMGENIPSTVSTISRNIRENVFKE 1038
Db 783 -----KKSEKDTNSKVSKVPGKVRWTSR-----TRGRWKYSNDESE 822
QY 1039 ASSNNIVGSSSTNEVGSSINEIGSSBENTQAEIAGNRGPKL-----NAMLRLGLVLP 1091
Db 823 GSGSEKSSAAEEBEEKEESEAIIADDEPKCKGCLNHPHLEILLDCSDGYHTACLRP 882
QY 1092 EYKOSLPGSN-----CKHPEIKQEEVEVQVNTDPSLYLSDNLEOPM---GSSHAS 1143
Db 883 PL--MIIPDGMFWPCQOH--KLCEKLEEOLODLOVALKKKAERKRLVVGISIE 939
QY 1144 QVCSETPDLLDGEIKEDTSFA-----ENDIKE 1172
Db 940 IIPQEPDFSEDEKCKKSKANLLERRSTTRKCI SYRDEFDEAIDEALEDI 999
QY 1173 SSVFSKSVQKGLSRSPFTHHLAQYR--RGAKKLESSENLSEDEELPCFQHLFP 1231
Db 1000 ADG-----GGVGRGKDITIT-----GHRGKDITILDEERKENRPPQRAAAARRKKR 1047
QY 1232 GKVNIPSOSTRHSTVATECLSKNTEENLLSKNSLNDSCNQVILAKASOEHLSEETKC 1291
Db 1048 RRLMDLSDSN-----LDEESEDEPKISDGSQD-----EFVVSDBNPD 1086
QY 1292 SASLFSOCSELEDTANTQDPLIGSSKOMRQHSQGVG-----LSDKELVSDDEE 1346
Db 1087 ESE-----EDPPSNDSDTDFC--SRRLRHPSPRPMRQSRRLRRTKPKKYSDD--1134
QY 1347 RGTGLENNQEQGMDSNLGEAAGCSETSVSDCGLSSQSDILTTQORDTMOHNLK 1406
Db 1135 -----EEESESENERD-----SEDFSDDFS-----DDFVETRRRRRRN---1169
QY 1407 LQOEAMAEVLEQHGQSPNSYPSIISDSSALEDLNRPQSTSEKAVLTOKSSEYPI 1466
Db 1170 -QKQINQKESDGSQKSLR-----RGKIRRVHRRKRLSSSESESYLS 1214
QY 1467 QNPEG-----LSADKFEVSADSTSKNKEPGVERSPSKCPSLDDRWMYH-----1511
Db 1215 KNSDEDELAKESKSVRKGRSTDEYSEADEEBEEBEGKPS-----RKRLRIETDEE 1267
QY 1512 -SCSGSLQNRNYPQBELIKVVDVEEQOLBESGPHDLTETSYPRLQDLE-----GTPYLE 1565
Db 1268 ESCDNAHGDAQPARDSQPRVLPSEQUESTKK--PYRIESDE---EEDFENVGVGSPLDY 1322
QY 1566 SGISLFDSDPESDSEDRAPEARVGNIPTSALKVPOLKVAESAGSPAAAHTTDTAG- 1624
Db 1323 SLVDLPSTNGOS-----PGKAIENLIGKPTKESQTPK-----DNSTASLASNGTSGG 1371
QY 1625 --YNAMESVSREKPELTASTERNV 1647
Db 1372 QEAGAPEE---EDELRLVTDLVD 1392

RESULT 10

US-09-815-242-5816
; Sequence 5816, Application us/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5816
; LENGTH: 2478
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5816

Query Match 3.2%; Score 312; DB 10; Length 2478;
Best Local Similarity 16.7%; Pred. No. 1.1e-08;
Matches 336; Conservative 343; Mismatches 794; Indels 538; Gaps 76;

QY 100 EYANSYNFAKKENNSPEHLKDEYSIIQSMGYRRA-----KRLLOSEPENPSLQ 148
Db 399 EIGNNGNF--ASLKADQFKYEVTLPQGVTVVNSLTTTPNGNEDSTVLKMTVNYDQN 456
QY 149 ETSLSVOLSNLGTVRLTRKQIQPKT---SVYIELGSSDSEDTVN---KATYCSVGQD 202
Db 457 ANKVTETSGQVTARGTHTKEVLPFKSLKSYKVNANIDTPKNIDFNEKLTYTASDV 516
QY 203 ELAQITPOGTRD---EISLDSAKKAACEFSETDVTNTEHHQPS-----N 243
Db 517 VINNAOPEVTLTADPVSVAVEMKMDALQQOVNSOVDNSHYTTASIAEYNKCLKQADTILN 576
QY 244 NDLN--TTEKRAAERHPE---KYGSSVSNLHV-----EPCGTNTNTHASSLQHENS 288
Db 577 EDANHVKTANRASQADIDGLVTLQAAALIDNAAIAELDTKAQEKVTAQAQSKKVTQDEV 636
QY 289 SLLLTQDRMNVK-----AEPCKNSKQPLARSQHNRWAGSKETCNDRTPTST--EKKVDL 342
Db 637 AALVTK--INNDKNNAAIAEINKQTTAQGVTTTEKDNGLAVLEQ---DVITPTVKPQAKODI 691
QY 343 NADPLCERKEWNKQKLPCSENPRDT-----EDVPWITLASSIQKVNWFERS 389
Db 692 -IQAVTTRKQOIIRKSNASLODEKDVANDKICKETKAIKOIDAATTTNAQVEAI-----743
QY 390 DELLGSDSDHSGESESNAKVADVLDVLNEVDEYSGSSEKIDLLASDP---HEALICKSR 446
Db 744 -KTKAINDINQTTPATTAQAA---ALEEFDEV--VQAQIDQAPLNPDTTNEEVAEAIER 796
QY 447 VHS-----KSVESNTEDEKIFGKYTYRKASLPNLSHVTENLIICAFVTEPOIIIGERPLTN 500
Db 797 INAAKVSQVKAIEATTTAQDLERV--KNEEISKIENTDST-----QTKMDAYNEVQAAA 849
QY 501 KLRKRRRPTSGLHPEDFIKKADLAVOKTPEMINOGTNOTEQNGOVMTNTHSHGNKTKGD 560
Db 850 TARKAQNATVSNATNEEVAEADAAVDAQAQ---QGLHDIQVVKSKQEVADTKSKVLDKIN 906
QY 561 SIONKPNPNPIESLEKESAFKTKAEPT-SSSISNMELELNHNSKAPKNNKLR-----612
Db 907 AIGTQAKVKPAADTEVENAYNTRKQEIQNSASTTEBKQAAYTETDITKQKQARTNLDAAN 966
QY 613 -----RKSTRHIALELVVSRNLSPNCTELQIDSCSSSEEEKKK 653
Db 967 TNSDVTAKDNSTAAINQVOAATTTKSDAKAEIAQKASERKTAIEAMNDSTTEEQQAQK 1026

QY 851 ELDA-----QYLONTFKV--SKRQFALFSPNPGNAEEECATFSA-----HSGSLKKQS- 896
Db 831 LANAKQPKQOLANLHLANDAKQKQSPESQITQAPLVTDTVTINQKAQALDHAMELLRNSI 890
QY 897 --PKVTFECE-----QKEENO--CKNESNIKPQVOTNITAGFPVGVQKDKPVDN 941
Db 891 ADMQATLASDYHDATAQKQNDYNAVTAKNIIIN-----QTSPTMNPDEVNKRATTQVNN 946
QY 942 AKSIRGSRFCFLSSQFRNETG----LITPNKHGILLQNPYRPPFPFKISFVKTKCKKN 997
Db 947 TKVALGDENLAAKQKQANRLNQLDLNNAKQKQLOSLQIAQSSDIAAVNGHKQTAESLN 1006
QY 998 LLEEN-----FEHSMSPERMGH-----ENIPSTVSTISRNIRENVFKE-ASSSNTINEVG 1048
Db 1007 TAMGNLINAIAHQVQEQR--GNFINADTKQATYATVNEAEAMINKOTGQANQPEVE 1064
QY 1049 SSTNEVGSSINEIGSSDENITQAEGLNRGPKLNAMLRLGLVQPEVVK-----1095
Db 1065 QAITKVOTTLQAL-NGDHLQV-AKTNATQAIADLTSLNDPQKLTALKDQVTAATLVAVH 1122
QY 1096 -----QSLPG-----SNCKH-----PEIKQEEVEVQ-----1118
Db 1123 QIBQANTLNQAMHGLRESIQDQNAATKANSKYINEDQPE-QQNYDQVQAANSIINEQT 1180
QY 1119 -----TYNTDFSPYLISNLQOPMGSSHASQVSETP-|-----DDLDD 1156
Db 1181 ATLDNNAINQAATVTNT--TKAALHGDVRLQNDKDKAKQTVSGLAYLNNAQKHMEDTLID 1238
QY 1157 GE-----IKED-----TSPAENDIKESSAVE-----1177
Db 1239 SETTRIVAKQDLTEAQAALQDLMTLOOSTADKDATRASSAYVNAEPNKKQAVDEAVONAE 1298
QY 1178 -----SRSVOKGELSRSPSPHT--HLAQGYRGAKKLESSEENLSEDELPQFQHL 1229
Db 1299 SIILAGLNPTINKGNVSSATQAVTTSKNGLDGVERLAQDKQTAGNSLNHLDLQITPAQQA 1358
QY 1230 LFGKVNIPQSOSTRHTVATECLSKWTEENLLSKNSLDCSNQVILAKASOEHHLSEET 1289
Db 1359 LENQINN-----ATTRDVAEIIAQQAALNEAMKALKESIKDPQTEASSKFINEQDQAOKA 1415
QY 1290 KCSASLFSQCSLELDTANTQDPFLGSSKQMRHOSQGV-GLSDKELVSDDEERG 1348
Db 1416 -----YTOAVQAHKDLI--NKTDDPLVKSVJDQATQAVNDKAKNLHGQKLAQDKQRA 1467
QY 1349 TGL-----BENQOEQMSNLGEAASGCESETSVSDCGSLSSQSDILTTQOQRTMOHN 1403
Db 1468 TETLNNLSNLTPOQALENQINNAATRGEVAKLTE-AQALNQAMEAL---RNSIQ-- 1520
QY 1404 LIKLOQEMAELEAVLBQHGQSQNSYPSIISDSSALEDLRNPESQSEKAVLTSQKSSEY 1463
Db 1521 ----DOQOTEAGSKFINEDKPKQDAYQAAVQAHAKDL--INQTSNPTLDKA-----1564
QY 1464 PISONPEGLSADKFEYSADSSTSKNKEPVERSSPKSPKSLDDRWYMHSCSGSLQNRNYP 1523
Db 1565 QVEQLTQGVNQAKDNLHGQKLA-----DDKHAVTDLNQLNSLNP 1606
QY 1524 SOBELIKVVDVEEQOLESGPHD-----LTETSYLP-----RQDLEGPTYESGISLFS 1572
Db 1607 QROAL-----ESQINNAATRDEVAQKLAQALDQAMQALRNSIQDQOQTESGSKFIN 1659
QY 1573 DD-PESDPSDEAPESARVGNIPSTSSALKVPQLKVAESAQAPAAHTTDTAGYNAMES 1631
Db 1660 EDPKQKD-----AYQAA-----VQAHKOLINOTGNPTLDKQSQVEQLTOA 1698
QY 1632 VSREKPELTASTERNKRMVMVSGLTPEEFMLVYKFARKHHI---TLNLI-----TE 1682
Db 1699 VITAKNLHGDKLARDQOQAVTT-----VNALPNLNAHQOALTDAINAAPTRE 1749
QY 1683 ETTHVVMKTDAEFVCERTLK 1702
Db 1750 VAQHVTQATELDHAME-TLK 1768

RESULT 14
US-09-987-482-1
; Sequence 1, Application US/09987482
; Publication No. US20020184656A1
; GENERAL INFORMATION:
; APPLICANT: BHANDARI, POONAM
; APPLICANT: SHASHIDHARA, L.S.
; TITLE OF INVENTION: IN VIVO ASSAY SYSTEM FOR SCREENING AND VALIDATION OF
; TITLE OF INVENTION: DRUGS AND OTHER SUBSTANCES
; FILE REFERENCE: 056859-0134
; CURRENT APPLICATION NUMBER: US/09/987,482
; CURRENT FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2843
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-987-482-1

Query Match 3.1%; Score 299.5; DB 9; Length 2843;
Best Local Similarity 18.9%; Pred. No. 6.2e-08;
Matches 398; Conservative 270; Mismatches 728; Indels 709; Gaps 93;

QY 69 TKRSLOESTRFSQL--VEELLKIIICAFOLD-----TGLE-----YAN 103
Db 874 SKRGLQISTTAAQIAKVMEEVSAIHTSQEDRSSGSTTLLHCVTDERNALRRSSAAHTSN 933
QY 104 SYNFAKKENNSPEHLKDEVSIIOSMGYRNPRAKRLLOSEPENSLOETSLSVOLSNLCTVR 163
Db 934 TYNFTKSENSNR-----TCSMPY-----AKLEYKRSSNDLSNSVSSDGYGRGQM- 979
QY 164 TLRTKRIQOPKTSVVYIELGSDSSDENVKATYCVSGVDQELLOITPOGTRDEI-----216
Db 980 -----KPSI-----ESYSEDESK--FCVSGY-----PADLAHKIHANHMD 1015
QY 217 ----SLDSAKKAACEFSETDVTNTEHHOPSNNDLANTTEKRAAE---RHPEKYQGSVSNI 269
Db 1016 DNDGELDTPTNLSKYSDQEL-NSGRQSPSQNERWARPKHIIIDEIKQEQQRQSNQSTT 1074
QY 270 HVEPCGTNTHASLQHNSSLLTLDKRMNVEKAEFCNKSQKQGLARQSHNRWAGSKETCN 329
Db 1075 Y----PVYTESTDDKH-----LAFQPHFGQOECVSPYRSRGANGSETNR-VGSNHGIN 1122
QY 330 DRRTPTSTKKKDLNADP-----LCERKEWNKQKLPCE-----ENPRD-----366
Db 1123 QNVSQLCEDDYEDDKPTNYSERYSEEEQHEEERPTNYSIKYNEEKRHVDQPIDYSLK 1182
QY 367 -TEDVPWITLNSIOKVNWFERSD-----ELGSDSDHGESESNKAVADVLDVLE 418
Db 1183 YATDIP-----SSQKQSPFSKSSSGQSSKTEHMSSENTSTPSSNAKQ-----NQ 1230
QY 419 VDEYSGSEKIDLLASDPHEALICKSERVHKSVE-----SNED 458
Db 1231 LHPSSAQSR-----SQQPKAATCKVSSINQETIOTYCVEDPTICFSRCSSLSUSSAED 1285
QY 459 KIFGTYRKAKSLPNLSHVTE-NLIIGAP-----VTEPQIIQERPLTNLKKRRKPTSL 512
Db 1286 ELGCGNQTQEADSANTLQIAEIKELGTRSAEDPYSEVPVAVSQHPRT---KSRLOGSSL 1342
QY 513 HPEDFIKA-----DLAVOKTPEMINOQNTQEQNQVNMNITN 550
Db 1343 SSESARHKAVFEFSSGAKSPSKGAQTPKSPPEHYVQETPLMFSRCTSVSSLD-----1394
QY 551 SGHENKTGDSIQNEK-----NPNPIESLEKE---SAFKTKAEPISISSINMELE 597
Db 1395 -SFESRSIASSVQSPCCSGMWSGIISPSDLPSQGTMPPSRSKTPPPPPQTAQKREVP 1453
QY 598 LNIHNSKAPKNNLRKRSSTRIHALELVVSNLSPN-----CTELQID---SCSS-- 646
Db 1454 KN-----KAPTAE--KRESGPKOA-AVNAAVQVQLVDPDADTLLHFATETPDGFCSSSL 1506
QY 647 -----SEEIKKKYNO-----657

Db 1507 SALSLEPFIQKDELRLTTPVQENDNGNETESQPKESNENQKEAKETIDSEKOLLDD 1566
QY 658 -----MPVRSRNLQLMGEPATGAKK-----SNKP----- 684
Db 1567 SDDDDIELEECIISAMPTKSR-----KAKKPAOTASKLPPPVARKPSOLPVYKLLPSQ 1621
QY 685 NBQTSKRHDS-----DTPELAK-LTMAPGSFTKCSNTSELKEFVNPSLPREEKEKLETVK 739
Db 1622 NRLQPKHVSFTPGDDMPRYVCVGTPIINFSTATSLSDL----- 1660
QY 740 VSNNAEDKDLMLSGERYLQTERSVESSISLVPTGYTQES-----ISLLEVSTLGKAK 795
Db 1661 ---TIESPPNELAAGEVGRGGAQSSEFEKRDITPEGRSTDEAOGGKTSVITPELDDNK 1717
QY 796 TEPNKCVCQAAPENPKGLIH-----GCKDNRRNTEGEFK-----Y 831
Db 1718 AEGDILAEICINSAMPKCKSHKPRVKKIMDVOQOASASSAPNKQLDGKKKKTSPVK 1777
QY 832 PLUGEVNH-SRETSTEMESELDAQYLQNTFKVSKROSFALFSN-----PGNAEEECAT 884
Db 1778 PIPQTEYTRVRKNADSKNMLNAERFSDNKKQNLKNNKSKDFNDKLPNNEDRVGRS 1837
QY 885 FSAHS-----GSL-----KQSPKVTFE 902
Db 1838 FAFDSPHHYTPLEGTTPYCFSRNDSLSLDDDDVDSLSREKAELKAKENKESAKVTSH 1897
QY 903 CEQKEENGKGNESNIKPVQTVNITAGFPV-----GOKDKP-----VD 940
Db 1898 TELTSNOOSANKTOAIAKQPIINRQOPKPILOKQSTFPQSKDIPDRGAATDEKLQNAIE 1957
QY 941 NAK-CSIKGGSRFLCS-----SQFRNETGLIT-----PNKHGILLQNPYRI--- 980
Db 1958 NTPVCFSHNSLSLSDIDQENNKNEPIKETEPPOSQSEPKPQASGAPKSFHVEDT 2017
QY 981 PLFP-----IKSFVTKCKKNLLEENFEHEHSPERMG-----NE 1017
Db 2018 POCFSRNSLSLSDIDEDLLQBCISSAMPKKKPSRLKGDKNEKHS---PRNMGGLIGE 2074
QY 1018 NIPSTVITISRNNTRENVFKASSNINEVGSSTNEVGSINEIGSSDENIQAEIQRNRG 1077
Db 2075 DLTLDKDIQRPDSEHGLSPDSFDMKATQEGANSTVSLHQAAAA-----ACLSRQAS 2129
QY 1078 PKLNAMRL--GV-----LQPVYKQSLPGSKHPELKQYEYEVQTVNTDF--- 1124
Db 2130 SDSLSILSKSGISLSPFHUTPD--QEEKPTFSNKPRIKPEKSTLETKKIESKSG 2187
QY 1125 -----SPYLIDNLQPMGSSHASVCSETPDDLLDGEIKEDTSAE 1167
Db 2188 IKGKKVYKSLITCKVRSNSELISQMKQPLQANMPS-----ISRGRTHIPGVR 2237
QY 1168 NDIKESSAVFSKSVQ-KGELSRSPSPTHLAQYRGAKKLESSEENLSSEDEELPCF 1226
Db 2238 NSSSTSPVKKGKPLKTPASKPS---EGQTATSPRGAPKSVKSELSVAR----- 2287
QY 1227 QHLLFGKVNIPQS--TRHST---VATECLSKNTEENLLSLKSLNDCNQVILAKASQE 1282
Db 2288 QTSQIGGSKAPSGSGRDSTSPPAQOPLSRPIQS---PGRNISPGRNGI-----SPP 2339
QY 1283 HHLSEETKCSASFSSOCSELEDITANTQDPFLIGSSK-----QMRHQSSEQGVG 1334
Db 2340 NKLSQLPRTSS-----PSTASTKSS-----GSGKMSVTSPPGRQMSQNLTKQTG 2383
QY 1335 LS-DKELVSDDEERGTLGLENNO-----EQGMSDNLGEAASGCSETSVSEDCSLG 1385
Db 2384 LSKNASSIPRSESASKGLNQNMNGNANKKVELSRMST---KSSGSEDRS---ERPVL 2437
QY 1386 SSQSDILTQOQDTHOHLNLIKLOEMAEALVLEQHSQP-----SNSYPSIISDSALE 1440
Db 2438 VRQSTTFKEAPSLRKL-----EESASFSL--SPSSRPASTRQOQTPVLSPSL--- 2488
QY 1441 DLRNPEOSTSKAVLTQKSEYPISONP-----EGLSADKFEVSADSTSKNKEPCVER 1495

Db 2489 ----PDMSLTHSSVOAGGWKRLPPNLSPTIEYDGRPAKRHRDIARSHSESPSRLP-INR 2543
QY 1496 S-----SPSKCPSLDDRWYMHSCSCLQNR 1520
Db 2544 SGTWKREHSHSSSLPRVSTWRRRTGSSSSILSASSESEKAKSEDEK-HVNSISGTQSK 2602
QY 1521 -NYPQBELIKVVDVEE-----QQLEESQPHDLTETSYLPQDLQEGTPYLESGISLFS 1572
Db 2603 ENQVSAKGTWRKTKENEFSTNSTQTVSSGATNGAESKTLIYQAPAVSKTEDVWVRIE 2662
QY 1573 DDPESDPSEDAPESARVGNIP---SSTSALKVPQLKVAESAQ----- 1612
Db 2663 DCPINPRSGRSP---TGTPPVIDSVSEKANPNIKDSQKQNVGNGSVPMRTVGL 2718
QY 1613 -----SFAAHTTDTAGN-----AMEESVSEKPELTASTERVNKRMSWV 1654
Db 2719 ENRLNSFIQVADAPDOKGTEIKPQNNPVPVSETNESSIVERTPFSSSSSKHSPSGTVA 2778
QY 1655 SGLTP 1659
Db 2779 ARVTP 2783

RESULT 15

US-08-681-219-32
; Sequence 32, Application US/08681219
; Patent No. US20020058607A1
; GENERAL INFORMATION:
; APPLICANT: Takaaki Sato and Junn Yanagisawa
; TITLE OF INVENTION: COMPOUNDS THAT INHIBIT THE INTERACTION BETWEEN
; TITLE OF INVENTION: SIGNAL-TRANSDUCING PROTEINS AND THE GLCF
; TITLE OF INVENTION: (PDZ/DHR) DOMAIN AND USES THEREOF
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/681,219
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/48962/JPW/JRM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-681-219-32

Query Match 3.1%; Score 298.5; DB 8; Length 2843;
Best Local Similarity 19.5%; Pred. No. 7e-08;
Matches 362; Conservative 275; Mismatches 662; Indels 561; Gaps 98;
QY 69 TKRSLOESTRFSOL--VEELLKIICAFQD-----TGLE-----YAN 103
Db 874 SKRGLQISTAAQIAKVMEEVSAIHTSQEDRSSGSTTELHCVTDERNALRRSSAAHTSN 933

Qy	104	SYNPAKKENNSP	HLKDEVIISIQSGMYRNRAKRLLQSPENPSLOETVSLSVOLSNLGTVR	163
Db	934	TYNFTKSENRR-----TCSMPY-----AKLEYKRSSNDLSNVSSSDGYGKRGQM-	979	
Qy	164	TLRTKRIQPKTSV	IELGSDSSEDTYNKATYCSVGQDELLOITPOCTRDEI-----	216
Db	980	-----KFSI-----ESYDDESK-----FCSYGOY-----PADLAHKTHSANHMD	1015	
Qy	217	----SIDSAAKAACEFSETDVNTTEHHOPNSNDDLNTTEKRAAE---RHPEKYOGSSVSNL	269	
Db	1016	DNGDELDTPI	NYSLKYSDBQL-NSGRQSPSONERWARPKHIIEDEIKOSEQRORNOSTT	1074
Qy	270	HVEPCGTHNASSLOHENSSLTTLKDRMNVKEAECNKSKOPGLARGOHNHRWAGSKETCN	329	
Db	1075	Y-----PVYTTESTDKH-----LKQPFGHQECVSPYRSGANGSETNR-VGSNHGIN	1122	
Qy	330	DRTPSTEKKVDLNADP-----LCEREKNWKOKLPCE-----ENPRD----	366	
Db	1123	QNVSOLCQODDYEDDKPNYSERYSEEQEHEEERPTNYSIKYNEKRHVDDQIDYSIL	1182	
Qy	367	-TEDVPWITLNSIQKWNEWFRSD-----ELLGSDSHDGESNAKVADVLDVLE	418	
Db	1183	KATDIP-----SSQKQSFPKSSGGOSSCTEHMSSESSENTSPSSNAKRQ-----NQ	1230	
Qy	419	VDEYSGSSEKIDLASDPHEALICKSERVHSKSVESN-IEDIKFQTYRKKASLPNLSHV	477	
Db	1231	LHPSSAQSR-----SQOPKAATCKVSSINOETIQTYCVEDTPI-CFSRCSSLSLSSA	1283	
Qy	478	TENLIICAFVTEPOIIQERPLTNKLKRKPRTSGLHPDEFIKKADLAVQKTPEMINOCTN	537	
Db	1284	EDE--IGC-----NOTTQEADSANTLO-----IAETIKEKI--GTR	1314	
Qy	538	QTQONGQVMNITNSGHENKTGDSTQNEKNPNPTBLESKESAFKTKAEPITSSSIINMELE	597	
Db	1315	SAED--PVSEVPVAVSOHPKTSRLQG-----SSLSESA-RHKAVEFSS-----	1356	
Qy	598	LNTHNSKAPKNKRLRR-KSTRH-----THALEVYVRNLSPNPNCOTELQID	642	
Db	1357	----GAKSPSKGAQPPKPPPEHYVOETPLMFSRCTSVSSLDSPESRSIA-----SSVOSE	1408	
Qy	643	SCS-----SSEEIKKKYNNOMPVRHS-----NLQLMGEKPATGAKKSNKP	684	
Db	1409	PCSGMVSGIISPDLSPGCQTMPPFSRSKTPPPPQTATQKREVPFNKAP-TAEKRESGP	1467	
Qy	685	N-----EOTSKRHDSTFPPELTLNAPSFTKCSNTSELKEFVNPSLPREEKEKLE	736	
Db	1468	KQAANVAQRVQVLPDATLLHFATESTPGDFS-CSSLSALSIDPEFFQKOVELRIMP	1526	
Qy	737	TVKVSNN-----AEDPKDMLSGERVLOTERSVSSSISLVPGTDYGTOESISLSLEVSTL	791	
Db	1527	PVQENDNGNETESEQPKSENQOEK-EAECTIDSEKOLL---DSDDDDDIEILEECII	1580	
Qy	792	GKAKTEPNKCVSOCA-AFENPKGLIHCSK-----DNR-----NDTEGFXYPL	833	
Db	1581	SAMPTKSSRKAKKPAQTASKLPPPVARKPSQLPVYKLLPSQNRLOPKQKHSYTFPGDDMPR	1640	
Qy	834	----GHEYNHSRETISIEMEESELDQYLONTFKVKRSQSFALFSPGN-ABEECATFS	886	
Db	1641	VYCVEGTPINFSTATSL---SDLIE-----SPNELAGEGVRRG	1678	
Qy	887	AHSGSLKKOSPKVTEFEQEKENOQKNESNIK-----PVQTVNITAGFPVVQG	934	
Db	1679	AQSGEFEKRD-TIPTEGRSTDQAQGGKTSVITIPELDDNKAEEGDILAECINSAMP-KGK	1736	
Qy	935	KDPVDNACKSTKGSRFLCSSQFRGNETGLITPNKHGILLONPRIPLPPIKSF-----	989	
Db	1737	SHTPPFVRVK-----IMDOVQOASASSAPKNOL--DGKHKKPTSPVRKPIPONTE	1784	
Qy	990	VTKTKCKNILLEEN-FEBHSMSPREMGENIPTSVTISRNNIRENVVENFEASSNINEV	1047	
Db	1785	YRTRVRKNADSKNNLNAERVFDNDKSKQNL-----KNSKD--FNDKLPNNEDRV	1834	
Qy	1048	-GS-----STNEVGSSINEIGSSDENITQAEELGNRPCKLANMLRLG	1087	

Search completed: January 22, 2003, 16:52:35
Job time : 77.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:44:08 ; Search time 34.5 Seconds
(without alignments)
5191.259 Million cell updates/sec

Title: us-09-734-672-4
Perfect score: 9649
Sequence: 1 MSLSALRVEEVONVINAMOK.....LYQCQLDYLIPQIPHSY 1863

Scoring table: BLOSOM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9642	99.9	1863	1 A58881	breast/ovarian can
2	4882	50.6	1812	2 I49350	breast/ovarian can
3	373	3.9	2447	2 T16870	hypothetical prote
4	365.5	3.8	3329	2 T42205	breast cancer susc
5	361.5	3.7	3343	2 T42207	breast cancer susc
6	358.5	3.7	5327	2 T13564	microtubule-associ
7	357.5	3.7	3329	2 T30904	breast cancer tumo
8	355	3.7	2954	2 T14156	kinesin-related pr
9	342	3.5	2271	2 F90073	hypothetical prote
10	341.5	3.5	1837	2 T41023	probable nuclear p
11	340.5	3.5	3187	2 JC8937	364K Golgi complex
12	340	3.5	2829	2 A42771	reticulocyte-bindi
13	339.5	3.5	3488	2 T34418	hypothetical prote
14	339	3.5	1957	2 T38077	hypothetical coile
15	334.5	3.5	3328	2 T30835	breast cancer tumo
16	333	3.5	6713	2 B89921	hypothetical prote
17	332	3.4	3225	2 I52300	giantin - human
18	330.5	3.4	3147	2 T18674	hypothetical prote
19	328.5	3.4	5170	2 T16241	hypothetical prote
20	327	3.4	1875	2 S38173	myosin-like protei
21	327	3.4	3418	1 G02334	breast cancer tumo
22	323	3.3	3259	1 A58539	giantin - human
23	322	3.3	1658	2 S55101	hypothetical prote
24	321.5	3.3	4377	2 A55575	ankyrin 3, long sp
25	320	3.3	1738	2 T14867	interaptin - slime
26	319	3.3	65	2 G02999	BRCA1 - rhesus mac
27	319	3.3	1410	1 A57013	early endosome ant
28	318.5	3.3	1326	1 A45605	mature-parasite-in
29	317.5	3.3	3507	2 T34513	hypothetical prote

Tpr homolog - frul
hypothetical prote
hypothetical prote
myosin-like coiled
fmbB protein (impo
probable heat choc
hypothetical prote
transport protein
cell proliferation
ankyrin 2, neurona
nuclear migration
adenomatous polyp
nuclear/mitotic ap
tanabair - African
probable myosin he
microtubule-associ

ALIGNMENTS

RESULT 1
A58881
breast/ovarian cancer susceptibility protein BRCA1 - human
C:Species: Homo sapiens (man)
C:Date: 16-Oct-1998 #sequence_revision 16-Oct-1998 #text_change 18-Jun-1999
C:Accession: A58881; A54652; I58130
R:SKolnick, M.H.
submitted to GenBank, September 1994
A:Description: Human breast and ovarian cancer susceptibility (BRCA1) mRNA, complete
A:Reference number: A58881
A:Accession: A58881
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1863 <SKO>
A:Cross-references: GB:U14680; NID:955931; PIDN:AAA73985.1; PID:955932
R:Wiki, Y.; Swensen, J.; Shattuck-Eidens, D.; Futreal, P.A.; Harshman, K.; Tavtigian, ye, C.; Hattler, T.; Phelps, R.; Haugen-Strano, A.; Katcher, H.; Yakumo, K.; Gholami, ow, P.K.; Norris, F.H.; Helvering, L.; Morrison, P.; Rostock, P.; Lai, M.; Barrett, J Science 266, 66-71, 1994
A:Authors: Lewis, C.; Neuhausen, S.; Cannon-Albright, L.; Goldgar, D.; Wiseman, R.; K A:Title: A strong candidate for the breast and ovarian cancer susceptibility gene BRC A:Reference number: A54652; MUID:95025896; PMID:7545954
A:Accession: A54652
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-444, 'D', 446-487, 'S', 489-593, 595-600, 'M', 601-1452, 'VLQ', 1456-1471, 'X', 14 A:Cross-references: GB:U14680
R:Hosking, L.; Trowsdale, J.; Nicolai, H.; Solomon, E.; Foulkes, W.; Stamp, G.; Signe Nature Genet. 9, 343-344, 1995
A:Title: A somatic BRCA1 mutation in an ovarian tumour.
A:Reference number: I58130; MUID:95315980; PMID:7795636
A:Accession: I58130
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1812-1813, 'PGQRTMASMQLGRCVRLHW' <HOS>
A:Cross-references: GB:S78558; NID:g1000568; PIDN:AAB34725.1; PID:g1000569
A:Note: frame shift mutation from an ovarian tumor
C:Genetics:
A:Gene: GDB:BRCA1
A:Cross-references: GDB:l36611; OMIM:113705
A:Map position: 17q21-17q21
C:Superfamily: breast/ovarian cancer susceptibility protein BRCA1; RING finger homolo
C:Keywords: mammary gland; nucleus; ovary; tumor suppressor; zinc finger
F:20-70/Domain: RING finger homology <RNG>
F:24-64/Region: zinc finger C3HC4 motif
F:607-621/Region: nuclear location signal
F:651-664/Region: nuclear location signal
F:679-692/Region: nuclear location signal
F:24,27,44,47/Binding site: zinc (Cys) #status predicted
F:39,41,61,64/Binding site: zinc (Cys, His, Cys) #status predicted
Query Match 99.9% Score 9642; DB 1; Length 1863;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MDLSALRVEEVQVNIAMOKILECPICLLEIKPEVSTKCDHIFCKMCLKLLNQKGPQ	60
DB	1	MDLSALRVEEVQVNIAMOKILECPICLLEIKPEVSTKCDHIFCKMCLKLLNQKGPQ	60
QY	61	CPLCKNDITKRSLOESTRFSQVVEELKIIKAFQDGTGLGLEYANSYFAKKENNSPEHLKD	120
DB	61	CPLCKNDITKRSLOESTRFSQVVEELKIIKAFQDGTGLGLEYANSYFAKKENNSPEHLKD	120
QY	121	EVSIIQSMGYRNRAKRLLOSEPENPSLOETSLSVQLSNLGTVRTLTQRIOQPKTSVYI	180
DB	121	EVSIIQSMGYRNRAKRLLOSEPENPSLOETSLSVQLSNLGTVRTLTQRIOQPKTSVYI	180
QY	181	ELGSDSSEDVFNKATYCSVGDDQELLQITPQGTTRDETSLDSAKKAACEFSETDVTNTEHQ	240
DB	181	ELGSDSSEDVFNKATYCSVGDDQELLQITPQGTTRDETSLDSAKKAACEFSETDVTNTEHQ	240
QY	241	PSNNDLNTTEKRAAERHPEKYQGSVSNLHVPCGTNTHASSLOHENSLLLTDKRMNVE	300
DB	241	PSNNDLNTTEKRAAERHPEKYQGSVSNLHVPCGTNTHASSLOHENSLLLTDKRMNVE	300
QY	301	KAFCKNSKQPLARQHRNAGSKETCNDRRTPSTEKKVDLNADPLCERKEWNKQKLPC	360
DB	301	KAFCKNSKQPLARQHRNAGSKETCNDRRTPSTEKKVDLNADPLCERKEWNKQKLPC	360
QY	361	SENPROTEDVPWITLNSSTOKYNEWFSRDELGSDSDHGESESNKAVADVLDVLEVD	420
DB	361	SENPROTEDVPWITLNSSTOKYNEWFSRDELGSDSDHGESESNKAVADVLDVLEVD	420
QY	421	EYSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKYRKRAKSLPNLSHVTEN	480
DB	421	EYSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKYRKRAKSLPNLSHVTEN	480
QY	481	LIIGAFVTEPQIIQERPLNKLKRRRPTSGLHPEDFTKKADLAVOKTPEMINQOTNOTE	540
DB	481	LIIGAFVTEPQIIQERPLNKLKRRRPTSGLHPEDFTKKADLAVOKTPEMINQOTNOTE	540
QY	541	QNGQVNIITNSGHENTKGDSTQNEKNPNPIESLKEKSAFTKAPPISSISNMLELENI	600
DB	541	QNGQVNIITNSGHENTKGDSTQNEKNPNPIESLKEKSAFTKAPPISSISNMLELENI	600
QY	601	HNSKAPKNLRRKSTRHIALELVVSRNLSPNCTELQIDCSSESBEIKKKYNQMPV	660
DB	601	HNSKAPKNLRRKSTRHIALELVVSRNLSPNCTELQIDCSSESBEIKKKYNQMPV	660
QY	661	RHSRNLQMEGKEPATGAKKSNKPNEQTSKRHDSDTFFPELKLITNAPGSFTKCSNTSELKE	720
DB	661	RHSRNLQMEGKEPATGAKKSNKPNEQTSKRHDSDTFFPELKLITNAPGSFTKCSNTSELKE	720
QY	721	FVNPSPREEKEKLETVKVSNNADPKDMLSGSERVLQTERSVESSISLVPDGYGTQ	780
DB	721	FVNPSPREEKEKLETVKVSNNADPKDMLSGSERVLQTERSVESSISLVPDGYGTQ	780
QY	781	ESISLLEVTGLKAKTEPNKVCYQCAAFENPKGLIHGCSKONRNDTEGFKYPLGHEVNH	840
DB	781	ESISLLEVTGLKAKTEPNKVCYQCAAFENPKGLIHGCSKONRNDTEGFKYPLGHEVNH	840
QY	841	RETSIEMESELDAQVQLQNTFKVSRKQSFALPSPNCGNAEEECATPSAHSGLKKOSPKVT	900
DB	841	RETSIEMESELDAQVQLQNTFKVSRKQSFALPSPNCGNAEEECATPSAHSGLKKOSPKVT	900
QY	901	FECEQKEEENOGKNEKSIKPVQTVNITAGFPVVGQDKDPVDNAKCSIKGSRFCLSQFRG	960
DB	901	FECEQKEEENOGKNEKSIKPVQTVNITAGFPVVGQDKDPVDNAKCSIKGSRFCLSQFRG	960
QY	961	NETGLITPNKHGLQNPYRIPPLFIKFSVTKCKKNLLEENFEHSHSPEREMGNENIP	1020
DB	961	NETGLITPNKHGLQNPYRIPPLFIKFSVTKCKKNLLEENFEHSHSPEREMGNENIP	1020
QY	1021	STVSTISRNNIRENVFKEASSSNIINEVGSSITNEIGSSDENIQAEIGRNRGPKL	1080

DB	1021	STVSTISRNNIRENVFKEASSSNIINEVGSSITNEIGSSDENIQAEIGRNRGPKL	1080
QY	1081	NAMLRLGVLPQEVYKQSLPGSNCKHPETKKQYEEVQVQNTDFSPYLLSDNLEQPMGSS	1140
DB	1081	NAMLRLGVLPQEVYKQSLPGSNCKHPETKKQYEEVQVQNTDFSPYLLSDNLEQPMGSS	1140
QY	1141	HASQVCSETPPDOLLDDGEIKEDTSAFENDIKESSAVFSKVGKGLSRSPSPFTHTHLAQ	1200
DB	1141	HASQVCSETPPDOLLDDGEIKEDTSAFENDIKESSAVFSKVGKGLSRSPSPFTHTHLAQ	1200
QY	1201	GYRRGAKKLESSEENLSEDEELPCFQHLLFKGVNNIPSQSTRHRSTVATECLSKNTEENL	1260
DB	1201	GYRRGAKKLESSEENLSEDEELPCFQHLLFKGVNNIPSQSTRHRSTVATECLSKNTEENL	1260
QY	1261	LSLKNLSLNDSCNOVTLAKASQEHLSSEETKCSASLFSQCSSELEDLTANTNTQDPLFSG	1320
DB	1261	LSLKNLSLNDSCNOVTLAKASQEHLSSEETKCSASLFSQCSSELEDLTANTNTQDPLFSG	1320
QY	1321	SKQMRHQSESQVGLSDKELVSDDEERGTGLEENQEEQSMDSNLGEAASGCESETSVSE	1380
DB	1321	SKQMRHQSESQVGLSDKELVSDDEERGTGLEENQEEQSMDSNLGEAASGCESETSVSE	1380
QY	1381	DCSGLSSQSDILTTQORDTMQHNLIKLOQEMAELEAVLEQHGQSPNSYPSIISDSSALE	1440
DB	1381	DCSGLSSQSDILTTQORDTMQHNLIKLOQEMAELEAVLEQHGQSPNSYPSIISDSSALE	1440
QY	1441	DLRNPQESTSEKAVLTQSKSEYPIQSONPEGLSADKFEVSADSSKKNKEPQGVERSSPSK	1500
DB	1441	DLRNPQESTSEKAVLTQSKSEYPIQSONPEGLSADKFEVSADSSKKNKEPQGVERSSPSK	1500
QY	1501	CPSLDDRMYMHSCSGSLQNRNYPQOEELIKVVDVEEQLEESGPHDLTSTSYLPRQDLEG	1560
DB	1501	CPSLDDRMYMHSCSGSLQNRNYPQOEELIKVVDVEEQLEESGPHDLTSTSYLPRQDLEG	1560
QY	1561	TPYLESGISLSDDDPESDRAPEARVGNIPSTSTALKVPQLKVAESAQSPAAAHHT	1620
DB	1561	TPYLESGISLSDDDPESDRAPEARVGNIPSTSTALKVPQLKVAESAQSPAAAHHT	1620
QY	1621	DTAGYNAMEESVSRKPELTASTERVNKRMSVMVSGLTPEEPMLVYKFKARKHITLNL	1680
DB	1621	DTAGYNAMEESVSRKPELTASTERVNKRMSVMVSGLTPEEPMLVYKFKARKHITLNL	1680
QY	1681	TEETHVVMYKTDAEFVCERTLKYFLGIAGGKVVSYFWVTQSIKERKMLNEHDFEVRGDV	1740
DB	1681	TEETHVVMYKTDAEFVCERTLKYFLGIAGGKVVSYFWVTQSIKERKMLNEHDFEVRGDV	1740
QY	1741	VNGRNHQGPKKARESDQRKIFRGLIICCYPPTNNPTDQLEHVMVQCGASVVKELSSFTL	1800
DB	1741	VNGRNHQGPKKARESDQRKIFRGLIICCYPPTNNPTDQLEHVMVQCGASVVKELSSFTL	1800
QY	1801	GTGVHPVIVVOPDAWTDNGFHAIGOMCEAPVVTREWVLDVALYOCQELDTYLIPOIPH	1860
DB	1801	GTGVHPVIVVOPDAWTDNGFHAIGOMCEAPVVTREWVLDVALYOCQELDTYLIPOIPH	1860
QY	1861	SHY 1863	
DB	1861	SHY 1863	

RESULT 2
I49350
breast/Ovarian cancer susceptibility protein BRCA1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 12-Nov-1999
C:Accession: I49350
R:Marquis, S.F.; Rajan, J.V.; Wynshaw-Boris, A.; Xu, J.; Yin, G.Y.; Abel, K.J.; Weber
Nature Genet. 11, 17-26, 1995
A:Title: The developmental pattern of Brcal expression implies a role in differentiat
A:Reference number: I49350; MUID:96021028; PMID:7550308
A:Accession: I49350
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1812 <RES>

A:Cross-references: EMBL:U31625; NID:g988213; PIDN:AAB17114.1; PID:g988214
C:Genetics:
A:Gene: Brcal
C:Superfamily: breast/ovarian cancer susceptibility protein BRCAL; RING finger homology
C:Keywords: zinc
F:20-70/Domain: RING finger homology <RNG>

Query Match 50.6%; Score 4882; DB 2; Length 1812;
Best Local Similarity 56.1%; Pred. No. 2.5e-209;
Matches 1046; Conservative 261; Mismatches 487; Indels 70; Gaps 32;

QY	1	MDLSALRVVEQVNVINAMQKILEPCICPLEIKPEVSTKCDHIFCKCMLKLLNKKGPSPQ	60
DB	1	MDLSAIOIEQVQVHMAQKILEPCICPLEIKPEVSTKCDHIFCKCMLKLLNKKGPSPQ	60
QY	61	CPUCKNDITKRSLOESTRESQVVEELKIIICAFOLDTGLEAYNSYFNAFKENNSPEHLKD	120
DB	61	CPUCKNEITKRSLOESTRESQVVEELKIIICAFOLDTGLEAYNSYFNAFKENNSPEHLKD	120
QY	121	EVSIIQSMGYNRRAKRLQSEPNPESLOETSLSVQLSNLGTVRLTKRIQIQPKTSVYI	180
DB	121	EASIIQSGVYRNVRRLPQVEPCNATLKD-SLGVQLSNLGIIVRSKKNRQIQPKTSVYI	179
QY	181	ELGSDSEEDTVNKATVCSVDQDQLLOITPOGTRDEISLSAKKAACEFSETDVTNTEHHQ	240
DB	180	ELDSDSSEETVTKPGDCSVRDQELLQTAQOAGDEGKLHSAEAAACEFSE-GIRNIEHHQ	238
QY	241	PSNNDLNTTEKRAAEHPEKQYSGSSVNLHVPCCGNTNTHASSILOHENSILLTKDRMNV	300
DB	239	CS-DDLNPTENHATERPEKQYSGSSVNLHVPCCGNTNTHASSILOHENSILLTKDRMNAE	297
QY	301	KAFCNKSQKQGLARSOHNRWAGSKETCDNRRTPTSEKKYVDLNLADPLCERKWNKQKLP	360
DB	298	KAFCNKSQKQGLAVSQSRAWAKGTCNDROVPGSTGEKVGPNADSLSDREKWTHTPQSLC	357
QY	361	SENPRATEDVPWITLNSQKQVNEVSRDELGLGSDSDHGESNAKVAADVLDVLENEVD	420
DB	358	PENSGATTDPWITLNSQKQVNEVSRTEGEMLTSDSARRHESNAEAAVLEVSNEVD	417
QY	421	EYSGSEKIDLLASDPHEALICKSEVHKSQSVESNTEDIKFKTYRKKASPLNLSHVTE	480
DB	418	GFSSSRKTDLVTPDHHITLCKSGDRDFKPVEDNLSKIFGKSYQKGRSRPHLNHVTE-	476
QY	481	LIIGAEVTEPQIIQERPLTNLKRKRRTSGLHPEDFIKKADIA-VQKTPEMINOQTNOT	539
DB	477	-IIGTEPTEPQITQEQPFTNKLKRRK--STSLQPEDFIKKADSAGVQRTPDNINQGTOLM	533
QY	540	EQNGQVNTNSGHENKTKGDSIQNEKNPNPIESLESKEFAFKTAEPISISSINMELELN	599
DB	534	EPNEQAVSTSNQENKIAGSNLQKESAHPTESLRKEPASTAGAKSINSVSQDLEVLN	593
QY	600	IHNSKAPKNRLRRKSTRIHALELVWSRNLSPPNCTELOIDCSSSEIEIKKKYNOMP	659
DB	594	VHSSKAPKNRLRRKSSIRCALPLE-PISENPSPTCAELQIDCSSSEIEIKKKYNOMP	652
QY	660	VHRNRLQMEGKEPATGAKSKNPNQTSKRHSDSTFPPELKITNAPGFTKCSNTSELK	719
DB	653	AGHLREPQIETDEPAADAKK-NEPNEHTKRRASDAFPEKLMNKAAGLLTSCSSPRKSQ	711
QY	720	EFVNPISLPREEKEKLETYKVSNNAEADPKDMLSGE-RVLQTERSVESSISILVPGTDY	778
DB	712	GPVNPSP-PORTGTQETQMSAKELGDRVLGGEPFGKTTDRSEESTSVSLVSDTDYD	770
QY	779	TQBSILLEVSTLIGKARTENKVCQAFENPKGLIHGCSKDNRRNDTEGFKYPLGHEVN	838
DB	771	TQNSVSLDAHTVRYARTGSAQCMQFVASENPKELVHG-SNNAGSGTEGLKPLRHALN	829
QY	839	HSRETSEMESELDQYQLNTQFKVSKROSFALESFNPNGNAEECAATFSAHSGSLKKQSPK	898
DB	830	LSQE-KVMEDESLDQYQLNTQFKVSKROSFALESFNPNGNAEECAATFSAHSGSLKKQSPK	884
QY	899	VTECEQKQENQCKNSNTKIPQVTYNITAGFPVVGQKDPVDNAKSIKGGSRFCUSSQF	958

RESULT 3
T16870

hypothetical protein T13H2.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Sep-2000

C:Accession: T16870

DB	885	VTAKGKQK-ERQOEFEFISHVOAATVGLPVPCEGKLAADTMD--RGCLCFSSH	941
QY	959	RGNETGLITPNKHLGLQNPYRIPPIKSFVTKCKKNLLEENFEHSHSPEREMNGEN	1018
DB	942	RSGENGLSATKSGISQNSHFQSVSPIRSIIKTDNRKPLTEGRFRHTSTTEWAVNGEN	1001
QY	1019	I-PSTVTSIRNNIRENVFKEASSNINEYSGSSTNEVGGSSINEIGSSINQIAELGNRG	1077
DB	1002	ILQSTVHTVSLNN-RGNACQEAQS-----GSIHEVCSTGDSFPQQLGNRG	1046
QY	1078	PKLNAMRLGLVLOPEYKQSLPCSNCKHPKIKOEVEEVVOTVNTDFSPYLISDNLEOPM	1137
DB	1047	PKYNTVPPLDMSMPGVCQQSVPSD-KYLEIKKOEBAVC----ADFSPLFSDHLEQSM	1101
QY	1138	GSSHASQVCSETPDDLLDDGEIKEDTSAFENDIKESAVFSKQVQKELSRSPSPETHH	1197
DB	1102	-SKVQVCSETPDDLLDDVEIOGHTSFGCEDIMERSAVFNGSILRRSSRSPTTHAS	1160
QY	1198	LAQGYRGAKKLESSSENJSEDEELPCFOHLLFGVNNIPQSSTRHSTVATECLSKNTE	1257
DB	1161	KSQSLHRASRKLESSESDSTEDLPCFOHLL-SRISNTP-ELTRCSSAVTQRMPEKAE	1218
QY	1258	ENLLSLKNSLDCSNQVILAKAQEHHLSEETKCSASLSSQCSLEDLTANTQDPFL	1317
DB	1219	GTQAPKMGSSDCNNEVITEASQEHQFSEDPRCSGSMFSSQHSAAQGSTANSDSNF	1278
QY	1318	IGSSKQMRHQSQSGVCLSDKELVSDDEBGTGLEENNOEQQSMDSNLCEA-ASGCESET	1376
DB	1279	IPSKQRSHOCGNEEAFSLDKELISNEEMATCLEBDNQEE--DSIIPDSASGYESET	1336
QY	1377	SVSEDCSGLSSQSDILTTQORDTMHNLIKLOEAMAELEAVLEQHGQSQSNYSPIISDS	1436
DB	1337	NLSEDC----SQSDILTTQRTATMKYNLIKLOEAMAELEAVLEQHGQSPSGHSPSLADP	1392
QY	1437	SALEDLRNPQSTSEKAVLTSQKSSEYPTSONPEGLSADKFEVS-ADSSTSKKNKPCVER	1495
DB	1393	CALEDLPLEPNMNSGAAILTSKNINENPVSNLKSACDDKFOQLHLEGPTSGDDESGMR	1452
QY	1496	SSPSKPSLDDRWYMHSCSGSLQNRNVPQOEELIKVVDVEEQOLEESGPHDLTETSYLPR	1555
DB	1453	PSFKPLAGSRGSAHCCSRHLQKRNPSQOEELLQPAGESE----ASPEHNSTGQSCCLR	1508
QY	1556	QDLEGTPYLESGLFSFD-DPESDPSSEDRAPESARVGNIPSSTSALKVPOLKYVAESAQSP	1614
DB	1509	RELEGTPYLGSGILFSRSDPSESPK---EPAHIGTTTASTSALKIPQGVAFRAAAA	1564
QY	1615	AAHTTDTAGYNAMESVSREKPELTASTERNVKNRMSWVSGLTPEEFMLVYFARKHHI	1674
DB	1565	AGAD-----KAVGVIVSKIKPELTSSSEERADRDISMVVSGLTPKVEMTVQKFAEYRL	1617
QY	1675	TLNLTETEETHVVMKTDABFVCERTLKVFLGIAGGKWVSYFWVTQSIKERMNLNEHDF	1734
DB	1618	TLTDAITEETHVIIKTDAEFVVCERTLKVFLGIAGGKWVSYFWVTQSIKERMNLNEHDF	1677
QY	1735	EVKGVVNGRNHGGPKRARSQDRKIFRGLIEICCYGFTTNMPTDQLEMMVQLGASVYKE	1794
DB	1678	EVKGVVTVGRNHGGPKRARSRE-KLFGKLVQVCCPEFTNMPKDELEMLQLCGASVYKE	1736
QY	1795	LSSTFLGTGVHPVIVVQPDWNTEDNGFHAIGOMCEAPVVTREMVLDVALYQCOELDTYL	1854
DB	1737	LPSLTHDTGAHLVIVQPSAWNTEDSNCPDGLQCKARLWMVDVLDSSLYRCRGLDAYL	1796
QY	1855	IPOI 1858	
DB	1797	VQNI 1800	

Db 1754 VSAPTKITAVPNVAAGFHQNOQQLYSMAQAQSTPIRTTTPQGTGSAPQAQTQSHLAQ 1813
QY 1590 VGNIPSTSAKVPOLK-----VAESAQSPAAAHTTDTAGYNAMESVSRE 1635
Db 1814 LGOFVNGANQQAPOQOQGMYTAAQLQMAQAAVAQTAQAQAAYAAE--AAYOAQVAQAARA 1872
QY 1636 KP 1637
Db 1873 AP 1874
RESULT 4
T42205
breast cancer susceptibility protein BRCA2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T42205
R:McAllister, K.A.; Haugen-Strano, A.; Hagevik, S.; Collins, N.K.; Brownlee, H.; Futreal
submitted to the EMBL Data Library, February 1997
A:Description: Characterization of the mouse and rat homologs of the BRCA2 breast cancer
A:Reference number: 222073
A:Accession: T42205
A>Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3329 <MCA>
A:Cross-references: EMBL:U89652; NID:g2443438; PID:g2443439; PIDN:AAB71377.1
A:Experimental source: strain CD1; 129Sv; ICR Swiss
C:Genetics:
A:Gene: BRCA2
C:Superfamily: breast cancer tumor suppressor BRCA2
Query Match 3.8%; Score 365.5; DB 2; Length 3329;
Best Local Similarity 19.4%; Pred. No. 5.1e-08;
Matches 402; Conservative 290; Mismatches 750; Indels 635; Gaps 93;
QY 72 SLQESTRFSOLVBEELKIIIC-----AFQDITGLEAYANSYNAFKKENNS---PEHLKD 120
Db 466 SLEDSTAGKQWVRSQAACLSPSIRKSIFKREPLDITLGTVFSDSMTNSTFTEEHEAS 525
QY 121 EVSIIQSMGYRNRAKRLQSEPNPS---LQETSLSVOLSNLGTVTTLTKQRIQPKT 176
Db 526 ACGILITACQBRDSICPSVSDTGSMTPTLTDTATVK--NAGLTSLTKNKKR----- 577
QY 177 SVYIELGSDSEDTVKATYCSVGDOBLQITPQGTREISLDSAKKAACEFSETDV--- 233
Db 578 -----KFIY-SVSDDASLOGKKLOTHROLELNL-SAQLAASAFVPLT 619
QY 234 -TYTEHHOPNNDLNITTEKRAAERHPKPYOGSSVSNLHVPCGTNTHASSLQHNSSL-- 290
Db 620 FTNVNSGIPDSSD-----KKRCLPNDE-----EPSLTNSFGTATSKEISVIHA 663
QY 291 LLTKDRMNYEKAFCNKSQKOPGLARQHNPNWAGSKETC-NDRTPST---EKKVDLNA-- 344
Db 664 LISQD-LNDKEAIVIEKPOPYTAREADFLLCPLPERTCENDQSPKVSNGKEKVLVSACL 722
QY 345 -----DPLCRKEWNKOKL-PCSENPRDTEVPW----- 372
Db 723 PSAVOLSSISFESQENPLGDHNGTSLKLTSPSKLPLSKADWVSREKMKMPKELQCESC 782
QY 373 ---ITLNSSTQKYNWFSSDELLGSDSDSHDGESENAKV-----ADVLVINEYDEY 422
Db 783 KVNIELSKNILEVNEICILSE-----NSKTPGLLPGCENIEVASSMKSQ 827
QY 423 SGSEKIDLLASDPHEALICKSR-----VHSKSVESNIEDKI-----FGKTYRKASLP 472
Db 828 FNQNAKI-----VIQKDQKGSFPTSEVAVNMNSEELFPDPSGNFNFAQVNTKCNKP 877
QY 473 NLSHVTENLIIGAVTEPQIIQERPL--TNKLRKRKRPSTGLHPEDFIKKAADLAVOKTPE 530
Db 878 DLGSSVE-----LQEDLSHTQGPGLSKLSPRAVDEVDAAHAQVLLITKDSD 924
QY 531 MINOGTNOTEQOVNMNITNSGHENKT-KGDSIQNEKNPNPNIESLEKESAFKTKAEP ISS 589

Db 925 SLAVVHDYTEKSRNNIEHQHQTEDKDFKSNSSLNMMKSDGNSDCSDKSEF---LDPVLN 981
QY 590 -----SISNMELENIHNSKAPK--NLRLRKSTR-----HIHALELVSRNLSP 634
Db 982 HNFGSPRTASNKEIKUSEHNVRKSMFFKDIEEQYPTRLACIDIVNTLPLANOQKILSEP 1041
QY 635 NCTELQIDSCSSSEIEKKKYNQMPVRHSRNLQLMCKEKPATGAKKSNKPNQETSKRHDS 694
Db 1042 HIPDLK-----SVTVSTQSHNQSSVSH-----EDT 1067
QY 695 DTPPELKLTNAPGSFTKCSNNTSELKEFPNPSLPREEKEKLETVK-----YSNNAEDP 747
Db 1068 DTAPQM-LSSKQDFHNNLTTSQAEITELSTILESGSQFETQFRKPSHIAQNTSEVP 1126
QY 748 KDLMLSGERVLIQT-ERSVESSSISLVPCTDYGQTESISLLEVLSTLGLKAKTPEPKVCSQA 806
Db 1127 GNMV-----VLSTASKEMKTDLHLVPDPSVGOTDHSKQFEGSA-GVKQSFPHLLEDTCN 1181
QY 807 AFENPKGLIHGCKSDNRNDTE--GFKYPLGHEVNHRSRETSTIEMEESLDAQYLQNTFKVS 864
Db 1182 --KNT-----SCLPNIEMEFEGFCALGTKLSVSE-----A 1213
QY 865 KROSFALFSPGNABEECA-----TFSAHSGS-----LKKQSPKVTFE---CEQK 906
Db 1214 LRKAMKLFSDIENSEEPSAKVGRGFSSSAHDSVASVFEKIKQNTBKSDFDEKSKQCVT 1273
QY 907 BENQ-----GKN-ESNIKPVQ-----TVNITAGFPVV---GQK 935
Db 1274 LONNIEMTTCIFVGRNPEKIKNTKHEDSYTSSQORNLENSDGSMSTSG-PVYIHKGDS 1332
QY 936 DKPVD-NAKCSIKGSRFCLSSOFRGNETGLITPNKHGLLQNPYRIPPLPIKSFVKTKC 994
Db 1333 DLADQSKC-----PESC--TQVARENTQIKENISDLTCLLEIMKAEETCMKSKDKQL 1385
QY 995 KKNLLEENFEHSMSPERMGNNIPSTVSTISRNIRENVFVREASSNINEVSGSTNEY 1054
Db 1386 PSDKMEONIKEFNIS-----FQTASGKNTR--VSRESLNKSVNIENRETDE- 1429
QY 1055 GSSINEIGSSDENIQAEQLGRNRPKLNAMLRLGVQPEVYKQSLPGSNCKHPKEL-KQEY 1113
Db 1430 ---LTVISDS-----LNSKILHGINKDKMH-----TSCHKKAISIKKVF 1465
QY 1114 EEVYQTVNTDFSP-----YLISDNLQPMQSGSHAS-----QVCSETPD---DILLDGE- 1158
Db 1466 EDHFPITVVSQLPAAQHPYEIESTKEPTLLSHTASGKVKIMQESLDKVNLFDETQY 1525
QY 1159 IKEDTSPAEND--IKESSAVFSKVQKSGELSRSPPTHTHLAQGYRRGAKKLESSENL 1216
Db 1526 VRKTASFQSGSKPLKSKELTLAYEKIEVT-----ASKCEEMQNFV 1567
QY 1217 SSEDEELPCFQHLLFGKVNNIQPSQTRHSTVATE-----CLSKN---TEENL 1260
Db 1568 SKETEMPLPOQNYHMYRQTEENLKTSGTSSKVQENIENNVEKNPRICICOSSYPVTDSA 1627
QY 1261 LS--LKNSLNDCSNQVILAK-----TAN-----TNTQ-----DPFLIGSK-- 1322
Db 1628 LAYTTEDSRKTCVRESSLSKGRKWLREQDKLGRNTWIKIECVKEHTEDFAGNASEHSL 1687
QY 1285 -----LSEETKCSASLFSQSCSELEDL----- 1306
Db 1688 VIIRTEIDTNHVSQVSTLLSDPNVCHVLSQSSFCCHDMHNDSGYFLKNKIDSDVPP 1747
QY 1307 -----TAN-----TNTQ-----DPFLIGSK-- 1322
Db 1748 DMRNAEGNTISPRVSATKERNLHPQITINEYCVQKLETNTSPHANKDVAIDPFLIDSRNCK 1807
QY 1323 -----QMRHQSQGVGLSDKELVSDDEERGTCLEBNNQEOESMDSNLGEAAGCSEET 1376
Db 1808 VGSIVFTAHSQETE----RTKEIVTDNCYKIV-----EONRQSKPDTCOTSCHKVL 1855
QY 1377 SVSEDCSGLSSQSDILTTQOQRTM-QHNLIKLOQEMAEVLEFQHGSGPSNSYPSTISD 1435

Db 1856 DSKDFICPSSGVDGVCINSRKDSFCPHNEQILQHNS-----MFLGKKAATPPVGLT 1908
Qy 1436 SSALEDLRNPEQ-----STSEKAVLTISOKSSSEYPISONPEGLSADKEEVSADS 1483
Db 1909 WDTSKSIRPEPPQAAHPSRYGTFSTASGKAIQVSDASLE-KARQVSEMDGDAKQLSSMV 1967
Qy 1484 STSKNKEP--GVRSSPSKCPSLDDRWMHSCSGSLQ-----NRNYP-----1523
Db 1968 SLEGNEKPHHSVKRENS-----VVHSTQGVLSPKPLPGNVNSSFSGFSTAGGK 2017
Qy 1524 ----SOEELIKVVDV-----EBQOLEESG-PHDLTFTSYLPQDLBGTYP-LESGIS 1569
Db 2018 LVTVSEALHKGMLLEFDLIRTEHTLQHSPIEDVSKILPQPCAEIRTPYVNSKLO 2077
Qy 1570 LFDDEPESDPRAPESARVGNIPSTGALKVPOLKVAESAQSPAAHTTDTAG-----1624
Db 2078 KTYNDKSSLSPNYK--ESSSGNTQSIIEVLSQLSOMERNQDTQLVLGTVKSHSKANLLGK 2135
Qy 1625 YNAMESVSREKPEL-TASTERVYNKRMVMVSGLTPEEFMLVIKFKARKHHTLTNLITEE 1683
Db 2136 EQTLPNQIKVKTDEMTFSDVPVKTWVGEYSKESENYFETEAVERSAKAFMEDDELTDSE 2195
Qy 1684 TTHVMKTDAEFCERTLYFIAGGKVVVYFWVTQSIKERKMLNEHD--FEVRGDVV 1741
Db 2196 QTHAKCSL---FTCPQNETLFNSRTRKRGVTVDAVGQPIKRSLNEDFDRIIESKGSKL 2252
Qy 1742 NGRNHQGPRARES---QDRKIF--RGLEICCYGPF 1772
Db 2253 T-----PSKSTPDGVTKDRSLFTHHMSLEPVCGPF 2283

RESULT 5
T42207
breast cancer susceptibility protein BRCA2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T42207
R:McAllister, K.A.; Haugen-Strano, A.; Hagevik, S.; Collins, N.K.; Brownlee, H.; Futreal
submitted to the EMBL Data Library, February 1997
A:Description: Characterization of the mouse and rat homologs of the BRCA2 breast cancer
A:Reference number: 222073
A:Accession: T42207
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3343 <MCA>
A:Cross-references: EMBL:U89653; NID:Q2443440; PID:Q2443441; PIDN:AA871378.1
A:Experimental source: strain Sprague-Dawley; testes
C:Genetics:
A:Gene: BRCA2
C:Superfamily: breast cancer tumor suppressor BRCA2

Query Match 3.7%; Score 361.5; DB 2; Length 3343;
Best Local Similarity 19.2%; Pred.No. 7.7e-08;
Matches 379; Conservative 297; Mismatches 715; Indels 579; Gaps 96;

Qy 66 NDIKTRSLQESTRFSOLVBELKLIICAFOLDGTGLYANSYNFAKKENNSPEHL-KDEYSI 124
Db 1087 HNLTPSQKAEITELSVILE-----ESGQFEFTQFN--PSHIAQNTSA 1129
Qy 125 IOMGYNRNRAKLQSEPNPNSLQTSLSVOLNLGTVRTLRTKORIOPKTSYVIELGS 184
Db 1130 V--LGNMAVVRTASEWKVDLH---LPLNPSSVGQIDHNKKFCELVGVKQS-----SS 1179
Qy 185 DSSEDVTKATCSVGDQELL---QITQGTDEISLSAKKAACEFSETDVTNTE-----237
Db 1180 HLLDTCNNTSCFLPIKEMFEGGFCGALGTVKLSVSNALRAMKLFSDIENISEEPSTK 1239
Qy 238 -----HH-----QPSNNDLNTTEKRA-----AERHPKY 261
Db 1240 VGRPFSSCAHDSVASVFKIKKQNTKDSFDEKSSKQVTVQNNKEMTTCILVDENPENY 1299
Qy 262 QGSSVNLHVEPCGNTTHASSLOHENSLLLTCKDRMYEKAFCNK--SKQGLARSOHN 319

Db 1300 ----VKNIQDNNTYTCSORNAYKLENSDV-----SKSSTSGTVYINKGDSDLF-FAAEKGN 1350
Qy 320 RWAGSKETC---NDRRTPTSEKKVDL-----NADPLCERKEWNKQKLPSCSENPRTDTE 369
Db 1351 KY---PESTQTVVRENAQIKESVSDLTCLVYMKAEETCHMKSSDKEQLPSDKMEQNMK 1407
Qy 370 VWITLNSIQVNEWFSRSDDELLGSDSHDGESNAKVADVLDVLNEVDEYSGSSEKI 429
Db 1408 ----FNLSFO-----TASGNIRVSKESLNKSVNILD--QETEDLTVTSDSL 1448
Qy 430 DLLASDPHEALICKSER-----VHSKSVESNIEDKIFOKTY--RKASLPLNLSHVTENL 481
Db 1449 ----NSKILCGINKDKMHISCHKKSI--NIK-KVFEHEFPIGTVSOLPALQOQPEYE 1498
Qy 482 IIGAFVTPP-----QIIQE-----RPLTNKLKRKRRTSGLHP-----E 515
Db 1499 I--ESIKEPTLLSFHTASGKKVKIMQESLDKVKNLFDETYVVKTTNFQHQESKPLKDRE 1556
Qy 516 DF-----IKKDLAVOKTPEMINQNTQTE-----QNG-----QVMNITNSGHENKTKGD 560
Db 1557 DYKERLTLAYEKIEVTASKCEEMQNFVSKQTEMLPQONDHMYROTENLTSNGSSPKVHG- 1615
Qy 561 STQN--ERNPNPIESLEKESAPKTKAEPIS-----SSISNMELELNINHSKA 605
Db 1616 NIENKIERNPR-ICCIQSSYFVETDSALACYTGDSTRKTCVGESSLSKGGKWLREQSDKL 1674
Qy 606 PKKNRLRKSSTRH-----IHALELVVSRNLSPPNCTELQIDSCSSSEELKKKYNO 657
Db 1675 GTRNTIEIQCKVEHTEDFAGNALYEHSLVIIRT-----EIDTSHVSENOAQLFSD 1725
Qy 658 MPVRHSRLQLMEGKEPATGAKKSNKPNEQTSKRHSDSTFPPELKLNTNAPGTSFTKCSNTSE 717
Db 1726 PNVCHS----YLSSHSCFCHDDMHNDSGYFLKDKIDSDVQPMKNKTEGNAIPKISATKE 1781
Qy 718 LKEFVNPLPREKEELETVKVSNNAEDPKDLMLSGERVLTQTERSVESSSISLVPCTDY 777
Db 1782 IK--LHPQTVNEECVQKLET-----NASPYANKNTAIDSAAMLDRNCKVGVSPVFI--TTH 1832
Qy 778 GTOESISLLEVTSLGKAK-----TERNKCVSQC-AAFENPKGLIHGCSKON--RNDYE 827
Db 1833 -SOETVRMKEIFTNCNKIVEONRESKPDTCQTSCHKALDNSEDFICPSSSGVDGVCINSPM 1891
Qy 828 GFKYPGLGHEV-NHSRETS-----IEMESELDAQYQLONTFKVKSRQSFALFS---873
Db 1892 AIFYQSQEILQHNQSVSLGKKAATPPVSVLEWTDCSTKIRGSPQEVHPHRTYGFSTASC 1951
Qy 874 ----NPGNAEECATFSAHSGSLKKQSPKVTFFCEQKEENOGKNESNI-----K 918
Db 1952 KAVQVSDASLEKARQVSEIDGDAKQLASVWSLEGNEKSHHSVKRESSVYVHNTHGVLSLR 2011
Qy 919 PVQTVNIT---AGFPVVQGDKPV-DNAKCSIKGSRFCLSSOPRGNETGLITPNKHGL 973
Db 2012 KTLPCNVSSFVSGFSTAGGLVTVSESALHKKVG-----MLEEFD-----LIRTEHL 2060
Qy 974 LQNP-----YRIPPLFKPFVKTKCKKNLLLENFEHSHMSPEREMGNENIPSTVSTIS 1028
Db 2061 QHSPTPEDVSKIPP-----QPCLESRTPEYSVS-----SKLQKTYNDKSR 2100
Qy 1029 NNIRENPFKEASSNINEVGSSTN-----EVGSSINIEIGSSDENIAELGRNRPKLNAML 1084
Db 2101 -----SPNSYKESGSGNTQSLVSPQLSQMERKQET-QSVLGTKVYSQRKTNIL 2148
Qy 1085 RLGVLPQEVYKQSLPGSNCKHPEIKKQVEEVVQTVNTDFSPYLLISDNLEQPMGSSHASQ 1144
Db 2149 E-----KKONLP-QNIKIESNKMETFSDVSMKTN-----VGE 2179
Qy 1145 VCSETPDLDDGETKEDTSEAFEND-IKESSAVFS-----SVQKGELSRSPSPFTHL 1198
Db 2180 YYSKEPENYFETEAVEIAKAFMEDDELTDSEQTHAKCSLFCACQNEALLNS-----RTRK 2234
Qy 1199 AQGYRGAKKLESSEENLSSEDEELPCFQHLHLLFGKVNNI-PSQSTRHSTVATECLSKNTE 1257
Db 2235 RGMAGVAVGQPPIKRSLNE-----FDRIIESKGSILTPSKSTPDGTTIKORRLF----2284

```
QY 1258 ENLLSLKNSLDCSNQVILAKASQEHLS-EETKCSASFSSQCELEDLTANTINTODPF 1316
Db 2285 -----THMSLEPVTGPF-----CSSKE-----RQETOSPH 2311
QY 1317 LIGSSQMR---HQSSQGVGLSD-----KELVSDEER-----GTGLENNQ 1356
Db 2312 VTPAQGLQKSHPSRHSAVYKSSSNPTVSALRSERTRHSVDSKSTKVFVPPFKVSRFH 2371
QY 1357 EQSMDS-NLGEAASGCESETSYSED-----CSGLSSQSDILTTQORDTMQH 1402
Db 2372 RDEHFSKNVLEGGKQKSDAGVSEGDNDSDFPQFNKDLMSLQNRADLDIIRKKERH 2431
QY 1403 NLKLOQMAEVALEVLEHQSQSPNSYPSTISDSSALEDLRNPEQ---STSEKAVLTSQ 1458
Db 2432 HLCPPQGLS-----YLTKSSTLPRISLQAAVGD--VPSACSPQLYMYGSKACISVNS 2484
QY 1459 KSEY---PISON--PEGLSADFEVSADSS---TSKNKEPGVGRSPSKC--PSLDD-- 1506
Db 2485 KNAEYFQFAIEDHFGKEALCAGKFRLAGDGGWLIPSDGKAGKEEFYRALLCDTPGVDPKL 2544
QY 1507 -----RWYHSCSG-----SLQNRNYPQOEELIKV--VDVEQQLLEESGPHD 1546
Db 2545 ISSWVSNHYRWIWLKLAEMFAFPKFPANRCLNPERVLQLKYRVDVEDNSRSALKK 2604
QY 1547 LTE-----TSYLPQDLEGTPLYLESGLSFLSDPSPDESDRAPESARVGNI----- 1593
Db 2605 ILERDDTAAKTLVLCVSDI-----ISLSTNVSETSGSKASSEDNKNVDITIELTDGW 2655
QY 1594 -----PSSTSALKVP--OLKVAESAQSP 1614
Db 2656 YAVKAQDPPLLALVKSGRLTVGQKIITOGAELVGPDPACAPLEAPDSLRLKISANSTR 2715
QY 1615 AAAHT-----TDTAGYNAMEEVSREKP---ELTASTERV--NKR 1649
Db 2716 ARWHSKLGFFHDPRLPPLSLSSFLSGGNGVCDVIVQRYVPLQWKEVYSGSYIFRNR 2775
QY 1650 MSVMVSLPPEEPMVLYKFARKHHITLNLITEETHVVMKMTDAEFVCR 1699
Db 2776 -----EERKEALRFAEOAKKLEALFTK--VHTELKEHEEDIAQR 2813

RESULT 6
Tl3564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG.4984.1
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: Tl3564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17689
A:Accession: Tl3564
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A:Note: EG:4984.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 3.7%; Score 358.5; DB 2; Length 5327;
Best Local Similarity 17.9%; Pred. No. 28-07;
Matches 362; Conservative 308; Mismatches 822; Indels 525; Gaps 75;

QY 31 IKEPVSTKCDHIFCKMLKLLNOKK-----GPSQCLCKNDITKRSIQES-TRF 79
Db 2355 VARTVSPPEEATMEFSKEVEVKESLALSLQGGSGCKLQTOSSPVDVAEGDFSHAVSV 2414
QY 80 SOLVEELLKIIACAFQDGTGLEVANSYNFAKKENNSPEHLKDEYSIIQSMGYRNRKRLIQ 139
```

```
Db 2415 STVTPLTKPAELAQGAAKTVSSPLDEALRTPSAPEHI-----SRADSPAE 2461
QY 140 SEPENSLQETSLSVOLSNLGTVRTLTQRIOQPKTSYVIELGSDSSDETVNKAICYSV 199
Db 2462 CASEELASQDKSPQVLKESRRPAWVAESKDDAAQLKSSVE-DLRSVPVASTEISRP--ASA 2518
QY 200 GDOELLIQTPQGTFRDEISLDSAKKAACEF-----SETDVTNTEHHQP--SNN 244
Db 2519 GETASSPI--EAPKDAEFAEQAEKAVLPLTIELKGNLPTLSSPVDVAHAS-VQPAELSKV 2576
QY 245 DLNTEKRAAERHPEKYQGSVSNLHVPCGTNTHASSLQHENSLLLTQRDNVNVKAEF 304
Db 2577 DIEKTASSPIDEAPKSLIGSPAERPEPAESAADAEE-----SVEKSK- 2620
QY 305 CNKSKOPGLARSQHNWAGSKETCNDRRTPSTEKKKVDLADPLCERKEWNKQKLPCESEN 364
Db 2621 -DASRPPSVVEST-----KADSTKGDISPSEVLGPKDDVEKSKESR--PPSVSA 2670
QY 365 RDTEDVPWITLNSIQVNEWFRSDELGSDSDSHGSESNAKVADVLDVLNEVEDEYSG 424
Db 2671 SITGD-----STKDVSRPASVSVESVDEHDKAESRRRESIAKVESVIDEACKSKSS 2722
QY 425 SSKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKAKSLPNLSHVNTENLIIG 484
Db 2723 SQDS---OKDEKSTLASKEARRRESVVESSKDDA-----EKSESRP----- 2760
QY 485 AFVTEPQIIQERPLTNKLRKRRPTSGLHPEDFIKKADLAVQKTPMINOGTNTQONGQ 544
Db 2761 -----ESVIAAGEVPVRESK-----SPLDSKDTSRFGSVVESYTAE-----DEKSEQSR 2805
QY 545 VMNITSGHENKTKGDSIQNEKNPNPIESLEKESAFKTAEPISSSISNMELNLHNHSK 604
Db 2806 RESVAESVADTKKCKSQEASRPSSVDELKDD--DEKQESRRQITGSHKAMTWGDE 2863
QY 605 APKKNLRKRSRTHIALELVVSRNL-----SPNCTELQIDSCSS--EEIKKKKYNQM 658
Db 2864 SPMDKADKSKEPSRPESVAESIKHENTKDBESPLGSR--DSVAESIKSDITKGEKSPL 2920
QY 659 PVHRSNLQMLECK--EPATGAKKSNKNEQTSKRHSDTFPELXLTNAPGFTKCSNT 715
Db 2921 PSKEVSRPESVSGSIKDEKAESRRRESVAESVKPESKSDATSAAPPKSHSRP----- 2971
QY 716 SELKEFVNPLSPREEKEKLETVKVSNNAEDPKDMLSGSERVLQTSERSVSSISLVPGT 775
Db 2972 -----ESVLGSLKDEGDKTTSRRVSVADSIKDEKSLVSOE-ASRPESEAEKDAAPSQ 3026
QY 776 DYTQESISLLEVTSLTGKAKTEPNKCVQCAAFENPKGLIHGCKDNRNNDTEGFKYPLGH 835
Db 3027 ETSRPESVT--ESVKGDKSPVASKEASRPASVAEN-----AKDSADES---KBQRP 3073
QY 836 EVNHSRETSTEESEELDAQ-YLQNTFKYVKRKQSF--LFSNPGNAEE----- 880
Db 3074 SLPQSRAGSIKDEKSLPLASKEAESKESRRRESVAESVQPLVSKESVSRPASVAESVKDEA 3133
QY 881 -----ECATFSAHSGSLKKQSPKVTPECKQKBNQKESNPKPVQTVNITAG 928
Db 3134 EKSKESPLMSKEASRPASVAGSVKDEAEK-----SKESRRESVAESKPLPSKEASRP 3187
QY 929 FVVGQDKPVDNAKCSIK---GGSRFLCSSQFRGNETGLITPNKHGLLQNPYRIPPLFP 985
Db 3188 ASVAESVKDEADKSKESRRRESGAESKPLASKEASRPAS----- 3226
QY 986 IKSFVTKCKCKNLLLE---ENFEHSMSPREMGNENIPSTVSTISRNNTRENVFKEASS 1042
Db 3227 VABSIKDEAEKSKESRRRESVAESKPLPSKEASR---PTSA-----KSVKDEAEKSKES 3279
QY 1043 NINEVSGSTNEVSGSSINEIGSSDENIQAEGLGRNRGPKLNAMLRLGLQ--PEVYKQ-SLP 1099
Db 3280 SRDSVAESKPLASKEASRPASVAESVQDDAEKSK-----EESRRESVAESKPLAYKEASRP 3335
QY 1100 GSNCK--HPEIKQOEYEEVVQTVNTDFSPYL-----ISDNLEQ----- 1135
```


Db 1214 LRKAMKLFSDIENSEEPAKVGPRGFSASAHSDVASVFKKIKQKQNTKESKDEKSSKQCVT 1273
QY 907 EENO-----GKN-ESNIKPVQ-----TVNITAGFPVV--GQK 935
Db 1274 LQNNIEMTTCIFVGRNPEKIKYKTHEDSYSSORNNLENSDGSMSSTSG-PVYIHKGDS 1332
QY 936 DKPVD-NAKSISIKGGRFCLSSQFRGNETGLITPNKHLGLQNPYRIPPLPIKSFVTKC 994
Db 1333 DLPADQSKC-----PESC--TOVARENTQIKENISDLTCLTEIMKAEETCMKSSDKQL 1385
QY 995 KKNLLENFHEHSMSPERGMENIPSTVSTISRNNIRENVFKREASSNNINEVSGSSINEV 1054
Db 1386 PSDKMEONIKFEFIS-----FOTASCKNTR--VSKESLNKSNIFNRFDE- 1429
QY 1055 GSSINETGSSDENITQABELGRNRPKLNALRLGVLOPEVYKQSLPGSNCKHPKIK-KOEY 1113
Db 1430 --LTVISDS-----LNSKILHGINKDKMH-----TSCHKKAISIKKVF 1465
QY 1114 EEVQVNTVDFSP-----YLISDNLEQPMGSSHAS-----QVCSETPD--DLLDDGE- 1158
Db 1466 EDHFPVIVTSQALPAQHPKYEIESTKEPTLLSPHTASGKKVKIMQESLDKVKNLFDETOY 1525
QY 1159 IKEDTFAEND--IKESSAVFSKSVQKGLSRSPSPHTHLAGQYRGAKKLESSENL 1216
Db 1526 VRKTASFQSGSKPLKDSKKELTAYEKIEVT-----ASKCEEMQNFV 1367
QY 1217 SDEDEELPCFQHLFLFGKVNIPQSOTRSTVATE-----CLSKN-----TEENL 1260
Db 1568 SKETEMPLQOYHYMYROTENLKTNSGTSSKQVENIENNVKFNPRICCIQOSSYPVTDISA 1627
QY 1261 LS--LKNSLNDCSNQVILAK-----ASOEHH- 1284
Db 1628 LAYITEDSRKTCVRESSLSRGRKWLREBQDKLGTNTIKIECVKEHTDFAGNASVEHSL 1687
QY 1285 -----LSKETCSASFSSQCSLEDL----- 1306
Db 1688 VIIRTEIDTHVSENQVSTLSDPNVCHLSYLSOSSFCCHDDMHNDGSYFLKKNKIDSDVPP 1747
QY 1307 -----TAN-----TNTQ-----DPFLIGSK-- 1322
Db 1748 DMKNAEGNTISPRVATKERNLHPQITNEYCVQKLENTSPHANKOVAIDPSLLDSRNCK 1807
QY 1323 -----QMRHQSSQGVGLSDKELVSDDEBRTGLENNQEOESMNLGAAAGCESET 1376
Db 1808 VGSLVFITAHSOETE-----RTKEIVTDNCKIV-----EQNRQSKPDTCOTCSCHKVL 1855
QY 1377 SVSDECSGLSQSDILTTQORDI-----MOHNLKLOQEMAELEAVLEHQSGQSPNS 1428
Db 1856 DSKDFICPSSGQDVCSNRKRSFCPHNEQILQHN-----QSMS-----GLKKAAT 1901
QY 1429 YPSTIISDSSALEDLRNPQ-----STSEKAVLTSQKSSEYPIQNPQEGLSADK 1476
Db 1902 PPVGLTWDTSKSIREFPQAAPHSPRYGIFSTASGKAIQVSDASLE-KARQVFSEMDGDA 1960
QY 1477 FEVSADSTSKNKEP--GVERSPSSKCPSLDWRMYHSGSGSLQ-----NRNYP----- 1523
Db 1961 KOLSSMVSLGNEKPHHSVRKENS-----VVHSTQGVLSLKPPLPGNVSSVFSG 2010
QY 1524 -----SQEELIKVDV-----EEQQLRESG-PHDLTETSYLPRLDLEGTPT 1563
Db 2011 FSTAGKLVTVSALHKGMLNEEDFLIRTEHTLQHSPIPEDVSKILPQPCAEIRTPY 2070
QY 1564 -LESGLISLFDSDPESDPRAPESARVGNIPSSSTALKVPOLKVAESAQSPAAHTTDT 1622
Db 2071 PVNSKLQTYNDKSLSPSNYK--ESSSGNTQSIQVLSQSMERNODTOLVLGTRKVSHS 2128
QY 1623 AG-----YNAMESVSREKPEL-TASTERVKNRMSVVGSLTPEEFMLVYTKFARKHITL 1676
Db 2129 KANLLGKEQTLQNIQKVTDEMKTFSQVVKTNVGEYSKESSENYFETEAVESAKAFMED 2188
QY 1677 TNLITEETHVVMKTDAEFCVCEFTLYFLGIAGKVVSVFWTQSIKERKMLNEHD--F 1734
Db 2189 DELTDEQTHAKCSL---FTCPQNETLFNSRTRKRGVTVDAVGOPPIKRSLLNEFDRII 2245

QY 1735 EVRGDVVNGRNHOGPKRARES---QDRKIF---RGLEICCYGPF 1772
Db 2246 ESKGKSLT-----PSKSTPDGTVKORSLETHHMSLEPVTGPF 2283

RESULT 8

T14156
Kinesin-related protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: T14156
R:Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.
Cell 91, 357-366, 1997
A:Title: CENP-E is a plus end-directed kinetochore motor required for metaphase chrom
A:Reference number: Z17893; MUID:98028574; PMID:9363944
A:Accession: T14156
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-2954 <WOO>
A:Cross-references: EMBL:AF027728; NID:g2586070; PID:g2586071; PIDN:AAC60300.1
C:Genetics:
A:Gene: XCENP-E
C:Superfamily: centromere protein E; kinesin motor domain homology

Query Match 3.7%; Score 355; DB 2; Length 2954;
Best Local Similarity 19.3%; Pred. No. 1.3e-07;
Matches 358; Conservative 276; Mismatches 678; Indels 546; Gaps 87;
QY 65 KNDITKRSLOESTRFVSOLVEELKIIICAFOLDTG-LEYANSYNPAKKE--NNSPEH---- 117
Db 552 EKETT--SLQQLQSKSE--BEKKELVQSFFELKIAELAEQLSVKAKNLEMTVNSREHSINA 607
QY 118 -----LKDEVSTLIQSMGYNRNRAKRLLOSEPENPSIQETSL----- 153
Db 608 EVQTDVEKEVVRKEMSVGLDSGY-----NASNSDLQSDSVGDKRLUSSSHDECIH 657
QY 154 -----VOLSNLGVVTLRTKRIQPKQTSVYIELGSDSSEDTVKNATYCSVGQELL 205
Db 658 RMLEQKIVDLEEF--FENLNKSKENDKQKS-----SEQDFMESIQLC-----EAI 701
QY 206 QITPQGTTRDISL--DSAKKAACEFSET---DVTNTEHHQPSNNDLNTTBKRAAERPEK 260
Db 702 MABKANALELALMRONFONILIE-NETLKREIADLERSLKENQETNEFEILEKETQKE- 759
QY 261 YQGSYSVNLHVEPCGINTHASSLOHENSLLLTKDRMNVKEAFECNKSQPG-----LA 314
Db 760 -----HEAQLIHEIGSL-----KKLVENAEYMNQNLLEEDLETTKLL 796
QY 315 RSQHNRWAGSKETCNDRRTPSTEKKV---DL-----NADPLCERKEWNKOKLPCSEN-PR 365
Db 797 KEQEIOLAEIR-----KRADNLQKVRNFDLSVSMGDSSEKLCEEIFOLKQSLSDAEAVTR 851
QY 366 DTED-----VPWITLNSSTQKYNWFSRDELLGSDSDSHDGSESNAK-----VADVL 413
Db 852 DAQKECSFLRSENLELKEKEMEDTSNNYNOKEAASLFEKOLETEKSNYKMEADLOKELQ 911
QY 414 DVLINEVDEYSGSSEKIDLLASDPHEALICKSE-----RVHSKSVESNIEDKIFGKYRKK 468
Db 912 SAFNEINYLNG-----LLAGKVPRLDLSRVELEKVKYSEFSKQLEKALEEK---NALENE 962
QY 469 AS-----LPN-----LSHVTENLIIGAFVTEPQIIQERPLTNKLRKRRTPTSGHL 513
Db 963 VTCLESEYKFLPNEVECLKNOISKASEIML-----LQKE-----GEH 999
QY 514 PEDFIKKADLAVOKTPEMINTQGTNQ-TEQNGQVYNTINSNGHENKTKGDSIQNEKNPNPIE 572
Db 1000 SASIISKQEIIMOQEQEIQILQTDVTHQTOSKVOQTEQYLEMKMKHDDL----- 1049
QY 573 SLEKESAFKTKAPISSSISNME-----LELNHNSKAPKKNRLRKRKSRTHIALELVV 627
Db 1050 -FEKYTRNKSEADLLREMENLKGTMESVEVKIADTKHELEETIROKEQLLH----- 1100


```
Db 775 EQSNHIVNDLKSNDLLTYSVKOKLADLNSLESKLSLQQDNFHMKAQIESNQYITATV 834
QY 662 HSRNLQIMEGKEPATGAKKSNPNEQTSKRHD-SDFPELKLNTNAPGSP-----T 710
Db 835 DSMNSRILE-----LSNDLRVANSKLSGSDVRLTLQN---SFDLREHQTILVQ 882
QY 711 KCSNTSELKEFVNPSLPR-----EKEEKLTVKVSNNAPDKDL 750
Db 883 LQSNITELQADI--TLQRTVRNQLEIQTTELKERLKFMEERQENLQSKLIAN---KDT 936
QY 751 MLSGERVQLTQERSVESSSISLVPGDYGTQESISLLEVSTIGKAKTEPKVCVQCAAFEN 810
Db 937 TQNPDDN-----EVEAISTEL-----ERTKEKLRAE---LEKSNIOQKYLASE----- 977
QY 811 PKGLIHGCKSDKNRDT-EGFKYPLGHEVNSRE--TSIE-----MEESLDA 854
Db 978 -KTL-----EMNNEHTEQFKHLVSEISTREKITSLSRSELLDNKRVEVLKEKESS 1030
QY 855 QYLQNTFKVSKROSFAFNSPNAEECATFSAGSKSLKQSPK---VTFECEQKEENQ- 910
Db 1031 KELAKQLEDAVREKDSALSFKDKYKIRSDADRVITSLKEDIKERSLMKECHSNEYSEI 1090
QY 911 ---GKNESNIKPVQT-----VNITAGFPV-----GOKD-----KPDVNAKCSIK 947
Db 1091 VSHGRTTQKRLDLRTFDEVNTKYLKLANFQOHSGLSGAEKDWNIOQKAMEDISSLK 1150
QY 948 G-----GSRECLSSQF-----RGNETGLITPNKHGILLQNPYRIPPLPIKSF 989
Db 1151 DYTILGLENQKLLHSQFDSLSQIITVLOONSSENINISANLEAVOONDLR-----ELVSI 1205
QY 990 VTKCKKNLLEENFEHSPEREMNENIPVSTISRNINRENVFKEASSN----- 1043
Db 1206 LRH-EKEIMDNKY-ELTILDNRGL-NOQVKSLOSTVDSLOLELNRLOSPLVSNQDTP 1261
QY 1044 -----INVEGSTNEVGSSINEIGSSDENIQIQLGRN 1075
Db 1262 IISGSOEVQLLYESNSVLKNDNAKLGKIQELEKEVEKELNASLNPLOTEINELKAEIGAK 1321
QY 1076 RG-----PKNLMLRLGLQVQYVYKQSLPGSNCKHPKPEIKOEYEEVVQ 1118
Db 1322 TASLNLKMEYNSRWKLRFOSVLNKNYERVDPTQLEBK-----NCEALEKEQOELTKLQ 1376
QY 1119 TVNTDFSPYLISNLEQPMGSSHASQVCSPTDPLDLDGEIKEDTSFAENDIKESAVFS 1178
Db 1377 ETAKE-----TDTFKQOVNS-----LNEEVENL-----KKEVEQANTKTRLAAWN 1418
QY 1179 ---KSVQKGLSRSPPFTHLAQYRGAKKLES---SEENLSSEDEELPCFOHLLFG 1232
Db 1419 EKCENLKSSLR-----FAHLKQELTNKNKELTSKNAENAMQKETEESLKDSNHQLOE 1472
QY 1233 KVNIPSQSTRHSTVATECLSKNTENLILSKNSLNDCSNOVILAKASQE-----HH 1284
Db 1473 SASSDAEQITKEQFOLKSEKERTKELADSKNELEHLOQSENAVADGKTEISNLEKEIHE 1532
QY 1285 LGEETKCSASLFSQCSSELDLTANTNTQDPP-----LIGSKQMRHOSQGV 1333
Db 1533 LRSDKGLVQOVQNLASAEALAREHSPHQSGLENADAETIARLSQLESTKQYYEREKETEI 1592
QY 1334 GLSDRELVSDDERTGTGLENNQOEQSDMSNGEAAACSETSVSDCGLSSQSDILT 1393
Db 1593 LAARSELVAEKEKTBELQENQNLNEKSQRRIKELEEQAOQNSSENT-----HDNIDDMIK 1645
QY 1394 TQORDTMOHNL---IKLOQEMAELE-----AVLEQHGQSPNSYPSIISDSSAL-ED 1441
Db 1646 QVVEKELKENSANFVKLKKVVAETEFPSKAKISVYERKTRDLQNKITQLEBETTENLAKQ 1705
QY 1442 LRNPEQS-----TSEKAVLTSQKSEPIYSONPEGLSADKFEVSADSSTSK-----NKE 1490
Db 1706 LSNPKTDESTSVTETKPTVTSKPTASKADVQGNATEASSAKREPSSGKLSARLOGTGKQ 1765
QY 1491 PGVERSSPK 1500
Db 1766 KGQRPVAVSR 1775
```

```
RESULT 11
JC5837
364K Golgi complex-associated protein - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
C:Accession: JC5837
R:Toki, C.; Fujiwara, T.; Sohma, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein
A:Reference number: JC5837; MUID:98093490; PMID:9431462
A:Accession: JC5837
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3187 <TOK>
A:Cross-references: DBJ:D25543; NID:g516825; PIDN:BAA05026.1; PID:g516826
C:Comment: This protein plays a role in the formation and maintenance of the characte
C:Superfamily: giantin
F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status pred
F:3165-3187/Domain: membrane anchor #status predicted <MAD>
```

```
Query Match 3.5%; Score 340.5; DB 2; Length 3187;
Best Local Similarity 18.9%; Pred. No. 6.2e-07;
Matches 375; Conservative 326; Mismatches 719; Indels 561; Gaps 91;
```

```
QY 23 ECPICLELIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKSLQESTRFSQL 82
Db 919 EEPVCKEALQOELE-----WLRKESEQRKKLQAAL---ISRKELIQ--KVSKL 962
QY 83 VEELKIKITCAFDLTGTGLEVANSYNFAKKENNSPEHLKDEVSIQSMGYRNRAKRLQSEP 142
Db 963 EEELAKV-----REESTKDSL-----RESEKRELEEDS 990
QY 143 ENP-----SLQETSLSVOLSNLGTVRTLRTKQRTQPOKTSVYIELGSDS 186
Db 991 KKKDDPEKYGTSEWRELEVSLRLTISEKEVELEGIKRRDLKEKAAAEELQALQVRM--- 1046
QY 187 SEDTVNKATYCSVGDOELL--QITPQ---GTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 1047 TQDLQNKTKQIDLQEEITENQATIQKFTGMDAGDGSQAVKETSVPSPPRAGGGEHWK 1106
QY 241 PSNDLNTTKRAAERHPKPYQGSVSNLHVPCGTNTHASSLQHSNLSLLLTDKRMNVE 300
Db 1107 PE-----LEKIVDLEKEKTQ-----LQKKLQEALISRKAI-LK 1139
QY 301 KAEFCNK-----SKQGLARSOHNRWAGSKETCNDRRTPSTPEKKVVDLNADPLCERKEN 354
Db 1140 KAQEKERHLKBEKQDAYRHLQEQFDGQSKENENIRAPLRQ-----LQAKEST 1189
QY 355 KOKLPCSENPRDTEVPWITLNSSIQKVNEWFSRDELLGSDSDSHDGES---ESNAKVAD 411
Db 1190 DQQLPGTGOQPTHGSEGLSEGT-----EPASESD-LHAAQPSHPGETATLQATVSVQAO 1243
QY 412 VLDVLNVEDEYSGSGEKIDLLASDPHEALICKSERV----- 447
Db 1244 IQDQLKELEV---EKEELEKISSTTSITSELTKKSEVLLLOEQINEQGLEIQNLKAASHEA 1300
QY 448 --HSKSVESNTEDKIFGTYYRKASLPNU-----SHVTE-----NLIIGAFVTE 489
Db 1301 KAHTEQKAOELE-----SSQLKIADLEHLKTLQPELETILQKHVGQKEEVEVSLVGOLGEK 1355
QY 490 PQITQERPLTNKLRKRPTSGLHP-----EDFIKKADLAV---QKTEMINQGTN 537
Db 1356 EOTL--TTVQTEMEEEQERLIKALHTQLEMQAKEERLUKQVQVEICELKKQKEEESK 1413
QY 538 QTEQNGQVNNITNSGHENKTKGDSIQNEKNPNPIESLEKE-SAFKTKABPISSSINMEL 596
Db 1414 AKQO-----LQRKLOAALISRKALKENKSLQEQLSARDADVEHLTKSLADVES 1462
QY 597 ELNTHNSKAPKKNRLRRKSSSTRHIIHALELVVSRNL-----PNCTELQIDSSCSSEE 649
Db 1463 QVSVQNG---EKDALLGKLALLQEEERDKLIVEMDKSLLENQSLGGSCESLKLALGLTID 1519
```


Query Match				3.5%; Score 334.5; DB 2; Length 3328;			
Best Local Similarity				18.9%; Pred. No. 1.2e-06;			
Matches 399; Conservative 286; Mismatches 724; Indels 701; Gaps 96;							
QY	72	SLOESTRFSQVLELLKIIIC-----AQOLDTGLYANSYFNAKKENNS----PEHLKD	120	QY	923	VNITAGPPVV---GQDKDPVD-NAKCSIKGGRFCLSSQFRGNETGLITPNKKHGLGNPY	978
Db	466	SLEDSIAGKQWRTSQAACLSPRSIRKSIFKMRPLEDLTGLTFVSDSMTNSTFTEEHEAS	525	Db	1318	MSSTSG-PVYIHKGDSLPADQSKC-----PESC--TQVAREENTQIKENISDLTCL	1369
QY	121	EVSTIQSMGYRNRAKRLIQSEBPNPS----LQETSLSVOLSNLGTVRTLTKORIQOQKT	176	QY	979	RIPPLPIKFSVTKCKKNLLEENFEHSPEREMGENIPSTVSTISRNNIRNVFXE	1038
Db	526	ACGLGILTACQREDSCIPSSVDTGSMPTTLTDTSATVK--NAGLISTLNKKR-----	577	Db	1370	MKAETCMKSDRKKQLPSPDKMERNIKEFNIS-----FOTASGKNTR--VSKE	1414
QY	177	SVYIELGSDSSEDTVKNATYCSVGDQELLQITPQGTREISLDSAKKAACEFSETDV---	233	QY	1039	ASSNSINEVSGSSINEIGSSDENIOAELGNRGPKNLMLRLGLVQLQPEVYKQSL	1098
Db	578	-----KFIY-SVSDASLQKKLQTHRQLELTNL-SAQLEASAFEVPLT	619	Db	1415	SLANKSVINFRDEL-----TVISDLSNKLHGINKDKMHTSCHKKALSIKKVFEDHF	1469
QY	234	TNTEHHQPSNNDLNTTEKRAAERHPEKYOSSVSNLHVEPCGTNTHASSLQHENSSL--	290	QY	1099	PGSNCKH-PEIKKOEYEEVVQTVNTDFSPYLIS-----DNLEQPMGSSHASQ	1144
Db	620	FTVNSGIPDSSD-----KKRCLPNDPE-----EPSLNSFGTATSKEISYVHA	663	Db	1470	PIVTVSOLPAQOQPEVE-----TESTKEPTLLSFHTASGKKVKIMOESLDK-----	1518
QY	291	LLTKDRMNVEKABFCNKSOPGLARSQHNRWAGSKETC-NDRRTTPSTEKKVDLNADPCE	349	QY	1145	VCSETPDDLDDGEIKEDETSFAEND--IKESSAVFSKVQKGLSRSPSPFTTHLAQGY	1202
Db	664	LISQD-LNDKEAIVIEEKPOPYTAREADFLCLPERTCENDQKSPKYSN-----	711	Db	1519	LCGETQ-----YVRKTASFQSGSKPLKDSKKELTAYEKIEVT-----	1556
QY	350	RKENKOKLPCSENPRDTEVPWITLNSSTQKYNEMFSRDELLGSDSDHDGESESNKV	409	QY	1203	RRGAKKLESSEENLSSEDELPCFQHLFGKVNNIPISQSTRHSTVATE-----	1250
Db	712	-----GKEKVLVS-----ACLPASVOLSSISFESQENHLG---DHNGTSTLKLTP	753	Db	1557	--ASKCEMNQNFVSKETEMLPQOONYHMYWOTENLKTSGTSKKVOENIENNYEKNPRIW	1613
QY	410	ADVLDVLNEVDEYSGSEKTDLLASDPHEALICK-SERVHKSVSIESNIEDKIFGKT----	464	QY	1251	CLSKN-----TEENLLS--LKNSLNDCSNOVILAK-----	1278
Db	754	SSKLPL-----SKADWYSREK-----MCKMPEKLQCESCKVNIE---LSKNILEV	795	Db	1614	CICQSSYPVTEDSALAYTTEDSRKTCVRESSLSKGRKWLREOGDKLGTRNTIKIECVKEH	1673
QY	465	-----YRKASLPNLSHVTENLIIGAFVTEPQIIQERPLNKLKRKRPTSGLHPDEFIK	519	QY	1279	-----ASQEHK-----LSEETHKCSASLFSQCSSELEDL----	1306
Db	796	NEICILSENSKTPGLLPPGNIETIVASSMKSQFNQNAKVIQDQKQSP-----FI-	846	Db	1674	TEDFAGNASYEHLVLIIRTEIDTNHVSNOVSTLLSDPNVCHSYLQSQSSFCCHDDMHDS	1733
QY	520	KADLAVOKTPEMI--NOGTNOTEQNGOVMNITNSGHE-----NKTKGDSIQN-----	564	QY	1307	-----TAN-----TNTQ-----	1313
Db	847	-SEVAVNMNSEELFPDGSNGFAQVTKCNKPDLGSSVELQEEDLSHTQGPSLKNSPMAV	905	Db	1734	GYFLKNKIDSDVPDMKNABEGNTISPRVSATKERNLHPQTINCYCVOKLETNTSPHANKD	1793
QY	565	-----EKNPNPIESLEK-----ESAPKTKA-----	584	QY	1314	---DPELIGSSK-----QMRHQSESGVGLSDKVELSDDEERTGTGLENNQEQSMD	1362
Db	906	DEDVDDAHAAQVILITKDSLSLAVVHDYTEKSRNNIEHQHQGTEDKDFKSNSSLNKMKS	965	Db	1794	VAIDPSLLDSRNCKVSGSLVFITAHSQETE-----RTKEIVTDNCYKIV-----	1841
QY	585	-----EPISS-----SIGNMELELNHNSKAPKK--NRLRKSSTR-----	618	QY	1363	SNLGEAASGCESETSVSDCSGLSSQSDILTTQOROT-----MOHNLKILQOEMAE	1414
Db	966	SDCSKXWSEFLDPVLNHNFCGSPRTASNKEIKLSEHNVKKSKMFFKDIEQYPTRLACID	1025	Db	1842	SKPDTQOTSCHVKLDDPSKDFICPSSSGDVCINSRKDSFCLHNEQILQHN-----	1894
QY	619	HIHALELWSRNLSPNPCTELQIDSCSSSEBEIKKKYNQMPVHRHSRLQLMEGKEPATGA	678	QY	1415	EAVLEHQGPSNPSYPSIISDSSALEDLNRPEO-----STSEKAVLTSQKSS	1462
Db	1026	IVNTLPLANGKKLSEPHIFDLK-----SVTTVSTQSHNQSSVSH-----	1064	Db	1895	-----GLKKAATPPVGLTWDTSKSIREFPQAAHPRTYIGIFSTASGKAIQVSDASLE	1947
QY	679	KSNKKNPQTSKRHSDTTPPELKLITNAPGSFTKCSNTSELKEFVNPSLPREEKKEKLETV	738	QY	1463	YPIQNPEGLSADKFEVSADSSTSKNKEP--GVERSPPSKCPSLDDRWYMHSCSGSLQ--	1518
Db	1065	-----EDTDTAPQM-LSSKQDPHSNNLTTSQKAEITELSTILEESGSOFEFT	1110	Db	1948	-KARQVFSEMDGDAKQLSSMVSLGEGNEKPHHSVKRENS-----VVHSTQGVLSLP	1996
QY	739	K-----VSNNAEDPKDMLSGERVLOT-ERSVESSISLVPGTGDYGTQESISLLEVST	790	QY	1519	-----NRNYP-----SQBELIKVVDV-----EQOOLEESG-PHDLTE	1549
Db	1111	QFRKPSHIAQNTSEVGNQW-----VLSTASKEWKOTDHLPLVDPDSVGQTDHKSQFEGSA	1166	Db	1997	KPLPGMNVNSSVFGSFTAGGKLVTVSESAHLHKVKGMLLEEDFLIRTEHTLQHSPIPEDVSK	2056
QY	791	LGKAETEPNKCVSQCAAFENPKGLIHGCSKDRNDRNTEGFKY--PLGHEVNHNSRETSTIEME	848	QY	1550	TSYLPQDQLEGTPY-LESGISLFDPSDPSDRAPEASARVGNIPSSTSALAKVPLKVA	1608
Db	1167	-GVKQSPFHLLDTCN-KNT-----SCFLPNINEMEFGRICSAIGTKLSVSNE-----	1212	Db	2057	ILPOPACAEIHTPEYPVNSKLOKTYNDKSLSPSNYK--ESGSSGNTOSIEVLSLOLSOMERN	2114
QY	849	ESELDQAQYQNTFKVKROGFALEFNPNGNAEECA-----TFSAHSGS-----LKKQ	895	QY	1609	ESAQSPAAAHTTDTAGYNAMESVSREKPELTASTERVNKRMSM-----VVSGLTPEEF	1662
Db	1213	-----ALRKAMKLFSDIENSEEPESAKVGRPGFSSSAHHDVSASVFKIKKP	1257	Db	2115	QDTQL-----VLGTVKSHSKANLKGKEQTLPKNIKVKTDEMKTFSDVPVKTN	2161
QY	896	SPKVTE-----CEQKEENQ-----GKN-ESNLIKPVQ-----	922	QY	1663	MLVYKFKARKHHIITLNLITEETHVVMK-----TDAE-----FVCERTPLKYFLGIAGG	1710
Db	1258	NTEKSFDEKSSKQVTLQNNIEMTTCIFVGRNPEKVIKNTKHEDSYTSSORNNLENSDGS	1317	Db	2162	VGEYYSKESENYFETEAV--ESAKAFMEDELDTSEQTHAKCSLFTCPQNETLFSNRTRK	2219
				QY	1711	KWVSYFWTQTSIKERKMLNEHD--FEVRGDVNGRNHQCPKARES---QDRKIF---R	1762
				Db	2220	RGVTVDAVGQPIKRIKSLLENEFDRIIESKGSALT-----PSKSTPDGTVKDRSLRLETHHM	2273
QY	1763	GLEICCYGPF	1772	QY			

Db 2274 SLEPVTGPF 2283

Search completed: January 22, 2003, 16:52:06
Job time : 96.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:43:17 ; Search time 17.5 Seconds
(without alignments)
4415.451 Million cell updates/sec

Title: US-09-734-672-4
Perfect score: 9649
Sequence: 1 MDLSALRVEEONVINAMOK.....LYQCQELDTYLIPQIPHSY 1863

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9642	99.9	1863	1 BRC1_HUMAN	P38398 homo sapien
2	6923	71.7	1878	1 BRC1_CANFA	Q95153 canis famli
3	4817	49.9	1812	1 BRC1_MOUSE	P48754 mus musculus
4	357.5	3.7	3329	1 BRC2_MOUSE	P97929 mus musculus
5	340	3.5	2869	1 RBP1_PLAVB	Q00798 schizosacch
6	339	3.5	1957	1 YD86_SCHPO	Q10411 schizosacch
7	327	3.4	1875	1 MLP1_YEAST	Q02455 saccharomyc
8	327	3.4	3418	1 BRC2_HUMAN	P51587 homo sapien
9	325	3.4	3911	1 AKA9_HUMAN	Q99996 h a-kinase
10	323.5	3.4	3210	1 CENF_HUMAN	Q99996 h a-kinase
11	322	3.3	1658	1 YM67_YEAST	P49454 homo sapien
12	321.5	3.3	4377	1 ANK3_HUMAN	Q03661 saccharomyc
13	319.5	3.3	2492	1 ATRX_HUMAN	Q12955 homo sapien
14	317.5	3.3	1972	1 P531_HUMAN	P46100 homo sapien
15	316	3.3	1727	1 ALM1_SCHPO	Q12888 homo sapien
16	313.5	3.2	1679	1 YI09_YEAST	Q99996 h a-kinase
17	312	3.2	3924	1 ANK2_HUMAN	P40457 saccharomyc
18	310.5	3.2	2845	1 APC_MOUSE	Q01484 homo sapien
19	308.5	3.2	2748	1 NUM1_YEAST	Q61315 mus musculus
20	308	3.2	2230	1 GOGA_HUMAN	Q00402 saccharomyc
21	307.5	3.2	1790	1 US01_YEAST	Q13439 homo sapien
22	304.5	3.2	1744	1 TANA_XENLA	P25386 saccharomyc
23	303.5	3.1	1539	1 Y373_HUMAN	Q01550 xenopus lae
24	301.5	3.1	2319	1 AKA6_HUMAN	Q15078 homo sapien
25	301	3.1	2464	1 MAPB_MOUSE	Q13023 homo sapien
26	299.5	3.1	2843	1 APC_HUMAN	P14873 mus musculus
27	296.5	3.1	1940	1 MYH3_HUMAN	P25054 homo sapien
28	294.5	3.1	2349	1 TPR_HUMAN	P11055 homo sapien
29	293	3.0	2468	1 MAPB_HUMAN	P12270 homo sapien
30	290	3.0	2663	1 CENF_HUMAN	P46821 homo sapien
31	290	3.0	3562	1 PGCV_CHICK	Q02224 homo sapien
32	288.5	3.0	1928	1 MYL1_YEAST	Q90953 gallus gall
33	288	3.0	2476	1 ATRX_MOUSE	P08964 saccharomyc
					Q61687 mus musculus

RESULT 1				
BRC1_HUMAN				
ID	BRC1_HUMAN	STANDARD;	PRT;	1863 AA.
AC	P38398;			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Breast cancer type 1 susceptibility protein.			
GN	BRCAL.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANT ARG-1775.			
RX	MEDLINE=95025896; PubMed=7545954;			
RA	Miki Y., Svendsen J., Shattuck-Eidens D., Futreal P.A., Harshman K.,			
RA	Tavtigian S., Liu Q., Cochran C., Bennett L.M., Ding W., Bell R.,			
RA	Rosenthal J., Hussey C., Tran T., McClure M., Frye C., Hattler T.,			
RA	Phelps R., Haugen-Strano A., Katcher H., Yakumo K., Gholami Z.,			
RA	Shaffer D., Stone S., Bayer S., Wray C., Bogden R., Dayananth P.,			
RA	Ward J., Tonin P., Narod S., Bristow P.K., Norris F.H., Helvering L.,			
RA	Morrison P., Rostock P., Lai M., Barrett J.C., Lewis C., Neuhausen S.,			
RA	Cannon-Albright L., Godlgar D., Wiseman R., Kamb A., Skolnick M.H.;			
RT	"A strong candidate for the breast and ovarian cancer susceptibility			
RT	gene BRCAL."			
RL	Science 266:66-71(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97092865; PubMed=8938427;			
RA	Smith T.M., Lee M.K., Szabo C.I., Jerome N., McEuen M., Taylor M.,			
RA	Hood L., King M.-C.;			
RT	"Complete genomic sequence and analysis of 117 kb of human DNA			
RT	containing the gene BRCAL."			
RL	Genome Res. 6:1029-1049(1996).			
RN	[3]			
RP	IDENTIFICATION OF BRCAL AS MEMBER OF BASC.			
RX	MEDLINE=20245492; PubMed=10783165;			
RA	Wang Y., Cortez D., Yazdi P., Neff N., Elledge S.J., Qin J.;			
RT	"BASC, a super complex of BRCAL-associated proteins involved in the			
RT	recognition and repair of aberrant DNA structures."			
RL	Genes Dev. 14:927-939(2000).			
RN	[4]			
RP	REVIEW ON VARIANTS.			
RX	MEDLINE=96400954; PubMed=8807330;			
RA	Couch F.J., Weber B.L.;			
RT	"Mutations and polymorphisms in the familial early-onset breast			
RT	cancer (BRCAL) gene." Breast Cancer Information Core.;			
RL	Hum. Mutat. 8:18-18(1996).			
RN	[5]			
RP	VARIANTS LEU-1637; GLU-1708 AND ARG-1775.			
RX	MEDLINE=95025878; PubMed=7939630;			
RA	Futreal P.A., Liu Q., Shattuck-Eidens D., Cochran C., Harshman K.,			
RA	Tavtigian S., Bennett L.M., Haugen-Strano A., Svendsen J., Miki Y.,			
RA	Eddington K., McClure M., Frye C., Weaver-Felhaus J., Ding W.,			
RA	Gholami Z., Soedervik P., Terry L., Jhanwar S., Berchuk A.,			

RA Iglehart J.D., Marks J., Ballinger D.G., Barrett J.C., Skolnick M.H.,
 RA "BRCA1 mutations in primary breast and ovarian carcinomas.";
 RT Science 266:120-122(1994).
 RL [6]
 RN
 RP VARIANTS BC GLY-64, AND VARIANTS ALA-772; ASN-1040 AND GLY-1443.
 RX MEDLINE=95201806; PubMed=7894491;
 RA Castilla L.H., Couch F.J., Erdos M.R., Hoskins K.F., Calzone K.,
 RA Garber J.E., Boyd J., Lubin M.B., Deshano M.L., Brody L.C.,
 RA Collins F.S., Weber B.L.;
 RT "Mutations in the BRCA1 gene in families with early-onset breast and
 RT ovarian cancer.";
 RT Nat. Genet. 8:387-391(1994).
 RL [7]
 RN
 RP VARIANT BC G-61, AND VARIANTS R-356; G-1038; N-1040; R-1183 AND
 RP G-1613.
 RX MEDLINE=95201808; PubMed=7894493;
 RA Friedman L.S., Ostermeyer E.A., Szabo C.I., Dowd P., Lynch E.D.,
 RA Rowell S.E., King M.-C.;
 RT "Confirmation of BRCA1 by analysis of germline mutations linked to
 RT breast and ovarian cancer in ten families.";
 RT Nat. Genet. 8:399-404(1994).
 RL [8]
 RN
 RP VARIANT BC GLY-61.
 RX MEDLINE=96108965; PubMed=8554067;
 RA Serova O., Montagna M., Torchard D., Narod S.A., Tonin P., Sylla B.,
 RA Lynch H.T., Feunteun J., Lenoir G.M.;
 RT "A high incidence of BRCA1 mutations in 20 breast-ovarian cancer
 RT families.";
 RT Am. J. Hum. Genet. 58:42-51(1996).
 RL [9]
 RN
 RP VARIANT BOC TRP-841.
 RX MEDLINE=97123469; PubMed=8968716;
 RA Barker D.F., Almeida E.F.A., Casey G., Fain P.R., Liao S.-Y.,
 RA Masunaka I., Noble B., Kurosaki T., Anton-Culver H.;
 RT "BRCA1 R841W: a strong candidate for a common mutation with moderate
 RT phenotype.";
 RT Genet. Epidemiol. 13:595-604(1996).
 RL [10]
 RN
 RP VARIANTS BC AND BOC.
 RX MEDLINE=96372821; PubMed=8776500;
 RA Durocher F., Shattuck-Eidens D., McClure M., Labrie F.,
 RA Skolnick M.H., Goldgar D.E., Simard J.;
 RT "Comparison of BRCA1 polymorphisms, rare sequence variants and/or
 RT missense mutations in unaffected and breast/ovarian cancer
 RT populations.";
 RL Hum. Mol. Genet. 5:835-842(1996).
 RL [11]
 RN
 RP VARIANTS BC MET-271 AND SER-1150.
 RX MEDLINE=96303704; PubMed=8723683;
 RA Katagiri T., Emi M., Ito I., Kobayashi K., Yoshimoto M., Iwase T.,
 RA Kasumi F., Miki Y., Skolnick M.H., Nakamura Y.;
 RT "Mutations in the BRCA1 gene in Japanese breast cancer patients.";
 RT Hum. Mutat. 7:334-339(1996).
 RL [12]
 RN
 RP VARIANT BC GLY-61, AND VARIANTS ARG-239; TRP-841 AND ILE-1512.
 RX MEDLINE=98430998; PubMed=9760198;
 RA Dong J., Chang-Claude J., Wu Y., Schumacher V., Debatin I., Tonin P.,
 RA Royer-Pokora B.;
 RT "A high proportion of mutations in the BRCA1 gene in German
 RT breast/ovarian cancer families with clustering of mutations in the 3'
 RT third of the gene.";
 RL Hum. Genet. 103:154-161(1998).
 RL [13]
 RN
 RP VARIANT BC GLY-64, AND VARIANTS.
 RX MEDLINE=98141685; PubMed=9482581;
 RA Andersen T.I., Eiken H.G., Couch F., Kaada G., Skrede M., Johnsen H.,
 RA Aloysius T.A., Tveit K.M., Tranebjærg L., Doerum A., Moeller P.,
 RA Weber B.L., Boerresen-Dale A.-L.;
 RT "Constant denaturant gel electrophoresis (CDGE) in BRCA1 mutation
 RT screening.";
 RL Hum. Mutat. 11:166-174(1998).
 RL [14]
 RN

RP VARIANTS BC S-22; L-461; D-465; V-552; S-892; D-960; I-1025 AND
 RP A-1047.
 RX MEDLINE=98272917; PubMed=9609997;
 RA Katagiri T., Kasumi F., Yoshimoto M., Nomizu T., Asaishi K., Abe R.,
 RA Tsuchiya A., Sugano M., Takai S., Yoneda M., Fukutomi T., Namba K.,
 RA Makita M., Okazaki H., Hirata K., Okazaki M., Furutsuma Y.,
 RA Morishita Y., Iino Y., Karino T., Ayabe H., Hara S., Kajiwara T.,
 RA Housa S., Shimizu T., Toda M., Yamazaki Y., Uchida T., Kunitomo K.,
 RA Sonoo H., Kurebayashi J.-I., Shimotsu K., Nakamura Y., Miki Y.;
 RT "High proportion of missense mutations of the BRCA1 and BRCA2 genes in
 RT Japanese breast cancer families.";
 RL J. Hum. Genet. 43:42-48(1998).
 RL [15]
 RN
 RP VARIANT OC ARG-1749.
 RX MEDLINE=20455732; PubMed=10486320;
 RA Gayther S.A., Russell P., Harrington P., Antoniou A.C., Easton D.F.,
 RA Ponder B.A.J.;
 RT "The contribution of germline BRCA1 and BRCA2 mutations to familial
 RT ovarian cancer: no evidence for other ovarian cancer-susceptibility
 RT genes.";
 RL Am. J. Hum. Genet. 65:1021-1029(1999).
 RL [16]
 RN
 RP VARIANT BC S-346, AND VARIANTS L-871; G-1038; R-1183 AND G-1613.
 RX MEDLINE=99254821; PubMed=10323242;
 RA Li S.-L., Tseng H.-M., Yang T.-P., Liu C.-H., Teng S.-J.,
 RA Huang H.-W., Chen L.-M., Kao H.-W., Chen J.H., Tseng J.-N., Chen A.,
 RA Hou M.-F., Huang T.-J., Chang H.-T., Mok K.-T., Tsai J.-H.;
 RT "Molecular characterization of germline mutations in the BRCA1 and
 RT BRCA2 genes from breast cancer families in Taiwan.";
 RL Hum. Genet. 104:201-204(1999).
 RL [17]
 RN
 RP VARIANTS OC, AND VARIANTS.
 RX MEDLINE=99214030; PubMed=10196379;
 RA Janezic S.A., Zlogas A., Krumroy L.M., Krasner M., Plummer S.J.,
 RA Cohen P., Glidea M., Barker D., Haile R., Casey G., Anton-Culver H.;
 RT "Germline BRCA1 alterations in a population-based series of ovarian
 RT cancer cases.";
 RL Hum. Mol. Genet. 8:889-897(1999).
 CC -!- FUNCTION: NOT KNOWN. MAY REGULATE GENE EXPRESSION INVOLVED IN
 CC -!- TRANSCRIPTIONAL REGULATION OF P21 IN RESPONSE TO DNA DAMAGE.
 CC -!- SUBUNIT: PART OF THE BRCA1-ASSOCIATED GENOME SURVEILLANCE COMPLEX
 CC (BASC), WHICH CONTAINS BRCA1, MSH2, MSH6, MLH1, ATM, BLM, PMS2 AND
 CC THE RAD50-MRE11-NBS1 PROTEIN COMPLEX. THIS ASSOCIATION COULD BE A
 CC DYNAMIC PROCESS CHANGING THROUGHOUT THE CELL CYCLE AND WITHIN
 CC SUBNUCLEAR DOMAINS. CTIP INTERACTS SPECIFICALLY WITH THE BRCT
 CC DOMAINS.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- POLYMORPHISM: THERE IS EVIDENCE THAT THE PRESENCE OF THE RARE FORM
 CC OF Q356R AND L871P POLYMORPHISMS MAY BE ASSOCIATED WITH AN
 CC INCREASED RISK FOR DEVELOPING OVARIAN CANCER.
 CC -!- DISEASE: BREAST CANCER (BC) IS AN EXTREMELY COMMON MALIGNANCY,
 CC AFFECTING ONE IN EIGHT WOMEN DURING THEIR LIFETIME. A POSITIVE
 CC FAMILY HISTORY HAS BEEN IDENTIFIED AS MAJOR CONTRIBUTOR TO RISK OF
 CC DEVELOPMENT OF THE DISEASE, AND THIS LINK IS STRIKING FOR EARLY-
 CC ONSET BREAST CANCER. MUTATIONS IN BRCA1 ARE THOUGHT TO BE
 CC RESPONSIBLE FOR 45% OF INHERITED BREAST CANCER AND MORE THAN 80%
 CC OF INHERITED BREAST AND OVARIAN CANCER (BOC). MOREOVER, BRCA1
 CC CARRIERS HAVE A 4-FOLD INCREASED RISK OF COLON CANCER, WHEREAS
 CC MALE CARRIERS FACE A 3-FOLD INCREASED RISK OF PROSTATE CANCER.
 CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC -!- SIMILARITY: CONTAINS 2 BRCT DOMAINS.
 CC -!- DATABASE: NAME=HotMolBase; NOTE=BRCA1 entry;
 CC WWW="http://bioinformatics.weizmann.ac.il/hotmolbase/entries/brca1.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U14680; AAA73985.1; -.

Query Match				99.9%; Score 9642; DB 1; Length 1863;			
Best Local Similarity				99.9%; Pred. No. 0;			
Matches 1862; Conservative				0; Mismatches 1; Indels 0; Gaps 0;			
QY	1	MDLSALRVEVQVNVINAMQKILECPTCLELIKPEVSTKDHIFCKFCMLKLLNKKQKGPSQ	60				
DB	1	MDLSALRVEVQVNVINAMQKILECPTCLELIKPEVSTKDHIFCKFCMLKLLNKKQKGPSQ	60				
QY	61	CPICKNDITKRSLOESTRFSOLVEELLKIICAFQLODTGLEAYANSYFPAKKENNSPEHLKD	120				
DB	61	CPICKNDITKRSLOESTRFSOLVEELLKIICAFQLODTGLEAYANSYFPAKKENNSPEHLKD	120				
QY	121	EVSIIQSMGYRNRAKRLLOSEPENPSLOETSLSVQLSNLGTVRTLTKRIQPKQTSVYI	180				
DB	121	EVSIIQSMGYRNRAKRLLOSEPENPSLOETSLSVQLSNLGTVRTLTKRIQPKQTSVYI	180				
QY	181	ELGSDSEEDTVNKATYCSVGDDQLLOITPQGTDRDEISLSAKKAACEFSETDVTNTEHHQ	240				
DB	181	ELGSDSEEDTVNKATYCSVGDDQLLOITPQGTDRDEISLSAKKAACEFSETDVTNTEHHQ	240				
QY	241	PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGTNTHASSLQHENSLLLTCKDRMNYE	300				
DB	241	PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGTNTHASSLQHENSLLLTCKDRMNYE	300				
QY	301	KAFCNKSQKOPGLARSOHNRWAGSKETCNDRRTPSTPEKKVLDNADPLCERKEWNKOKLPC	360				
DB	301	KAFCNKSQKOPGLARSOHNRWAGSKETCNDRRTPSTPEKKVLDNADPLCERKEWNKOKLPC	360				
QY	361	SENPROTEVPWITLANSQKYNWFVSRDELGSDSDHGDGSESNAKVADVLDVLNEVD	420				
DB	361	SENPROTEVPWITLANSQKYNWFVSRDELGSDSDHGDGSESNAKVADVLDVLNEVD	420				
QY	421	EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN	480				
DB	421	EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN	480				
QY	481	LIITGAVTEPQIIQERPLTNLKRKRRTPSGLHPEDFIKKADLAVOKTPEMINQGTNOTE	540				
DB	481	LIITGAVTEPQIIQERPLTNLKRKRRTPSGLHPEDFIKKADLAVOKTPEMINQGTNOTE	540				
QY	541	QNGOVNNTNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPDISSISNMELELNI	600				
DB	541	QNGOVNNTNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPDISSISNMELELNI	600				
QY	601	HNSKAPKKRLRRKSTRHITHALELVVSNLSPNCTELOIDSCSSSEBIKKKKYNOMPV	660				
DB	601	HNSKAPKKRLRRKSTRHITHALELVVSNLSPNCTELOIDSCSSSEBIKKKKYNOMPV	660				
QY	661	RHSRNLQMEGKEPATGAKSKNPNQTSKRHSDSDFPELKLTPNPGSFTKCSNTSELKE	720				
DB	661	RHSRNLQMEGKEPATGAKSKNPNQTSKRHSDSDFPELKLTPNPGSFTKCSNTSELKE	720				
QY	721	FYNPSLPREEKEKLETVKVSNNADPKDMLSGERVLOTERSVESSSIISLVPDGTGYTQ	780				
DB	721	FYNPSLPREEKEKLETVKVSNNADPKDMLSGERVLOTERSVESSSIISLVPDGTGYTQ	780				
QY	781	ESISLLEVTSLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDETEGPKYPLGHEVNH	840				
DB	781	ESISLLEVTSLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDETEGPKYPLGHEVNH	840				
QY	841	RETSIEMEESLDAQYLQNTFKVSKRQSFPALFSPNGNAEEECATFSAHSGSLKKQSPKVT	900				
DB	841	RETSIEMEESLDAQYLQNTFKVSKRQSFPALFSPNGNAEEECATFSAHSGSLKKQSPKVT	900				
QY	901	FECOEKEEQQGNESNIKPVQVNTITAGFPVVGQDKPDVNAKCSITKGSRCFLSSQFRG	960				
DB	901	FECOEKEEQQGNESNIKPVQVNTITAGFPVVGQDKPDVNAKCSITKGSRCFLSSQFRG	960				
QY	961	NETGLITPNKHGLLQNPYRIPPLFTKTSVKTKCKNLLNEENFEHSMSPEREMGNENIP	1020				
DB	961	NETGLITPNKHGLLQNPYRIPPLFTKTSVKTKCKNLLNEENFEHSMSPEREMGNENIP	1020				

QY	1021	STVSTISRNIRNENVEKASSNNINEVSGSTNEVGSSINEIGSSDENIQAEIGRNRGPKL	1080
DB	1021	STVSTISRNIRNENVEKASSNNINEVSGSTNEVGSSINEIGSSDENIQAEIGRNRGPKL	1080
QY	1081	NAMRLGVLOPEVYKOSLPCSNCKHPEIKKQYEEVVQVNTDFSPYLLISDNLEQPMGSS	1140
DB	1081	NAMRLGVLOPEVYKOSLPCSNCKHPEIKKQYEEVVQVNTDFSPYLLISDNLEQPMGSS	1140
QY	1141	HASQVCSETPDLLDDGEIKEDTSPAENDIKESSAVFESKVGKSGELSRSPSTHTHLAQ	1200
DB	1141	HASQVCSETPDLLDDGEIKEDTSPAENDIKESSAVFESKVGKSGELSRSPSTHTHLAQ	1200
QY	1201	GYRRGAKKLESSEENLSSDEELPCFQHLFGKVNNIPSQSTRHRSTVATECLSKNTEENL	1260
DB	1201	GYRRGAKKLESSEENLSSDEELPCFQHLFGKVNNIPSQSTRHRSTVATECLSKNTEENL	1260
QY	1261	LSLKNSLDCSNQVILAKASQEHHLSEETKCSASLFSOCSELEDLTANTNTQDPFLIGS	1320
DB	1261	LSLKNSLDCSNQVILAKASQEHHLSEETKCSASLFSOCSELEDLTANTNTQDPFLIGS	1320
QY	1321	SKOMRHQSEQGVGLSDKELVSDDEERGTLLENNQOEQSDSNLGEAASGESETSVSE	1380
DB	1321	SKOMRHQSEQGVGLSDKELVSDDEERGTLLENNQOEQSDSNLGEAASGESETSVSE	1380
QY	1381	DCSGLSSQSDIILTQORDTMQHNLIKLOQEMAELEAVLRQHGQSQPSNSYPSIISDSSALE	1440
DB	1381	DCSGLSSQSDIILTQORDTMQHNLIKLOQEMAELEAVLRQHGQSQPSNSYPSIISDSSALE	1440
QY	1441	DURNPQSTSEKAVLTQSOKSEYPISONPEGLSADKFEVSADSSSTSKNKEPGEVRSRSPK	1500
DB	1441	DURNPQSTSEKAVLTQSOKSEYPISONPEGLSADKFEVSADSSSTSKNKEPGEVRSRSPK	1500
QY	1501	CPSLDDRWYMHSCSGSLQNRNYPQSEELIKVVDVEEQLEESGPHDLTSTYLPDQDLGG	1560
DB	1501	CPSLDDRWYMHSCSGSLQNRNYPQSEELIKVVDVEEQLEESGPHDLTSTYLPDQDLGG	1560
QY	1561	TPYLESIGISLFDSDPESDRAPEARSARVGNIPSTSAKVPQLKVAESAQSPAAAHHT	1620
DB	1561	TPYLESIGISLFDSDPESDRAPEARSARVGNIPSTSAKVPQLKVAESAQSPAAAHHT	1620
QY	1621	DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFKARKHITLNL	1680
DB	1621	DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFKARKHITLNL	1680
QY	1681	TEETHVVMKTDAEFCERTLYFLGIAGGKVVSVFYFWTQSIKERRKMLNEHDFEVRGDV	1740
DB	1681	TEETHVVMKTDAEFCERTLYFLGIAGGKVVSVFYFWTQSIKERRKMLNEHDFEVRGDV	1740
QY	1741	VNGRHHQPKRAESQDRKIFRGLIECCYGPFTNMPTDQLEWVQLCGASVVKELSSFTL	1800
DB	1741	VNGRHHQPKRAESQDRKIFRGLIECCYGPFTNMPTDQLEWVQLCGASVVKELSSFTL	1800
QY	1801	GTGVHIPVYVQDAWTEGNGFHAIGQMCAPVVTREWLDLSVALYQCQELDYILPQIPH	1860
DB	1801	GTGVHIPVYVQDAWTEGNGFHAIGQMCAPVVTREWLDLSVALYQCQELDYILPQIPH	1860
QY	1861	SHY 1863	
DB	1861	SHY 1863	

RESULT 2
BRC1_CANFA
ID BRC1_CANFA STANDARD; PRT; 1878 AA.
AC Q95153;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Breast cancer type 1 susceptibility protein homolog.
GN BRCAL.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

Db	1437	IADSCSPEDLLNPEQNASER-VITSEKSDSPISQNPESLSTDKFQVFLDSTSKNGEPG	1495
QY	1493	VERSSPSKCPSLDRWMMHSCSLQNRNYPSEELIKVVDVEEQOLESGPHDLTETSY	1552
Db	1496	MIRSSPSOSRLDLRWYVHSCPSRLQDNCPSOKELTKVVSMEEQPTSEARDLMEQSY	1555
QY	1553	LPRDLEGTPYLSGSLFSDPESDPEDRAPESARVGNIPSTSAKVPQPKVAESAQ	1612
Db	1556	LSRPDLGAPYLSGSLFSDPESDPSSHRASELAVHSSMPTSTSAKLPQPOVEESAK	1615
QY	1613	SPAAHTTDTAGYNAMESVSRKPKDELTAETERNVRMSVVSGLTPPEPMLVYKARKH	1672
Db	1616	STAAHIASTAGYNKSGEDSVIGIEKPEVISSTGVNKRISVASGLTPKEFMLVHKFARKH	1675
QY	1673	HITLNLITEETHVVMKTKDAEVCERTKYFLGIAGGKVVVSFYVWVTOSIKERKMLNEH	1732
Db	1676	HISLNLISEETHVMKTKDAEVCERTKYFLGIAGGKVVVSFYVWVTOSIKERKMLDEH	1735
QY	1733	DFEVGDVNVGRNHQGPRA-----RESQDRKIFRGLICCCYGPFTNMPTDQLEWVQLC	1787
Db	1736	DFEVGDVNVGRNHQGPRAKRESQDRKIFRGLICCCYGPFTNMPTDQLEWVHLC	1795
QY	1788	GASVWKLSEFTLGTGVHPVTVVQPDWEDNGFHAIGQMCAPVVTREHWLDSVALYQC	1847
Db	1796	GASVWKEPSLFTLSKGTHPVVVQPDWEDSGFHAIGQMCAPVVTREHWLDSVALYQC	1855
QY	1848	QELDTYLIPOIPHS 1861	
Db	1856	QELDTYLIQIPRT 1869	
RESULT 3			
ID	BRCL_MOUSE	STANDARD;	PRT; 1812 AA.
AC	P48754; Q60957; Q60983;		
DT	01-FEB-1996 (Rel. 33, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Breast cancer type 1 susceptibility protein homolog.		
GN	BRCL.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6; TISSUE=Embryo;		
RX	MEDLINE=96177659; PubMed=8634697;		
RA	Abel K.J., xy J., Yin G.Y., Lyons R.H., Meisler M.H., Weber B.L.;		
RT	"Mouse Brcl: localization sequence analysis and identification of		
RT	evolutionarily conserved domains.";		
RL	Hum. Mol. Genet. 4:2265-2273(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6;		
RX	MEDLINE=96177660; PubMed=8634698;		
RA	Sharan S.K., Wims M., Bradley A.;		
RT	"Murine Brcl: sequence and significance for human missense		
RT	mutations.";		
RL	Hum. Mol. Genet. 4:2275-2278(1995).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=129/SVJ;		
RX	MEDLINE=96121367; PubMed=8575748;		
RA	Bennett L.M., Haugen-Strano A., Cochran C., Brownlee H.A.,		
RA	Fiedorek F.T. Jr., Wiseman R.W.;		
RT	"Isolation of the mouse homologue of BRCL and genetic mapping to		
RT	mouse chromosome 11.";		
RL	Genomics 29:576-581(1995).		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=129/SVJ; TISSUE=Embryo;		
RX	MEDLINE=96067162; PubMed=7590247;		

Lane T.F., Deng C., Elson A., Lyu M.S., Kozak C.A., Leder P.;

"Expression of Brcl is associated with terminal differentiation of

ectodermally and mesodermally derived tissues in mice.";

Genes Dev. 9:2712-2722(1995).

[5]

SEQUENCE OF 727-1111 FROM N.A.

STRAIN=C57BL/6; TISSUE=Embryo;

MEDLINE=96021028; PubMed=7550308;

RA Marguis S.T., Rajan J.V., Wynshaw-Boris A., Xu J., Yin G.Y.,

Abel K.J., Weber B.L., Chodosh L.A.;

"The developmental pattern of Brcl expression implies a role in

differentiation of the breast and other tissues.";

Nat. Genet. 11:17-26(1995).

[6]

SEQUENCE OF 789-1250 FROM N.A.

STRAIN=129/SVJ;

MEDLINE=96163506; PubMed=85666965;

RA Schroeck E., Badger P., Larson D., Erdos M., Wynshaw-Boris A.,

Ried T., Brody L.;

"The murine homolog of the human breast and ovarian cancer

susceptibility gene Brcl maps to mouse chromosome 11D.";

Hum. Genet. 97:256-259(1996).

RL

CC FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. INVOLVED IN

TRANSCRIPTIONAL REGULATION OF P21 IN RESPONSE TO DNA DAMAGE (BY

SIMILARITY).

CC SUBUNIT: CTIP INTERACTS SPECIFICALLY WITH THE BRCT DOMAINS (BY

SIMILARITY).

CC SUBCELLULAR LOCATION: Nuclear (Potential).

CC TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED IN OTIC VESICLES AT

DAY 9.5. AT DAY 10.5, THIS EXPRESSION DECREASES AND HIGH LEVELS

ARE FOUND IN THE NEUROECTODERM. AT DAYS 11-12.5, HIGH LEVELS IN

DIFFERENTIATING KERATOCYTES AND WHISKER PAD PRIMORDIA. AT DAYS

14-17, EXPRESSION ALSO OBSERVED IN KIDNEY EPITHELIAL CELLS. IN

THE ADULT, HIGHEST LEVELS FOUND IN SPLEEN, THYMUS, LYMPH NODES,

EPITHELIAL ORGANS, AND ALVEOLAR AND DUCTAL EPITHELIAL CELLS OF,

THE MAMMARY GLAND. VERY LOW LEVELS IN BRAIN, KIDNEY, AND SKIN. NO

EXPRESSION IN HEART, LIVER OR LUNG.

CC DEVELOPMENTAL STAGE: IN THE MAMMARY GLAND, EXPRESSION INCREASES

DRAMATICALLY DURING PREGNANCY. LEVELS FALL DURING LACTATION AND

INCREASE AGAIN DURING POST-LACTATIONAL REGRESSION OF THE

MAMMARY GLAND.

CC SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.

CC SIMILARITY: CONTAINS 2 BRCT DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; U35641; AAB17113.1;

EMBL; U31625; AAB17114.1;

EMBL; U32446; AAA96393.1;

EMBL; U36475; AAC52323.1;

EMBL; U33835; AAA99742.1;

MGD; MGI:104537; Brcl.

InterPro; IPR001357; BRCT.

InterPro; IPR002378; Brst_cancer1.

InterPro; IPR001841; Znf_ring.

Pfam; PF00097; zf-C3HC4; 1.

Pfam; PF00533; BRCT; 2.

PRINTS; PR00493; BRSTCANCER1.

SMART; SM00292; BRCT; 2.

SMART; SM00184; RING; 1.

PROSITE; PS50172; BRCT; 2.

PROSITE; PS00518; ZF_RING_1; 1.

PROSITE; PS50089; ZF_RING_2; 1.

Zinc-finger; DNA-binding; Nuclear protein; Anti-oncogene; Repeat.

ZN_FING 24 65 RING-TYPE.

DOMAIN 1585 1679 BRCT 1.

DOMAIN 1698 1797 BRCT 2.

```
FT VARIAT 93      F -> L (IN STRAIN 129/SVJ) .
FT VARIAT 305     T -> S (IN STRAIN 129/SVJ) .
FT VARIAT 319     Q -> A (IN STRAIN 129/SVJ) .
FT VARIAT 377     Q -> E (IN STRAIN 129/SVJ) .
FT VARIAT 550     K -> Q (IN STRAIN 129/SVJ) .
FT VARIAT 652     A -> P (IN STRAIN 129/SVJ) .
FT VARIAT 765     S -> P (IN STRAIN 129/SVJ) .
FT VARIAT 917     P -> L (IN STRAIN 129/SVJ) .
FT VARIAT 933     C -> S (IN STRAIN 129/SVJ) .
FT VARIAT 1122    K -> I (IN STRAIN 129/SVJ) .
FT VARIAT 1206    S -> R (IN STRAIN 129/SVJ) .
FT VARIAT 1212    RM -> GI (IN STRAIN 129/SVJ) .
FT VARIAT 1255    S -> R (IN STRAIN 129/SVJ) .
FT VARIAT 1261    H -> N (IN STRAIN 129/SVJ) .
FT VARIAT 1264    V -> A (IN STRAIN 129/SVJ) .
FT VARIAT 1269    P -> A (IN STRAIN 129/SVJ) .
FT VARIAT 1283    T -> K (IN STRAIN 129/SVJ) .
FT VARIAT 1337    T -> N (IN STRAIN 129/SVJ) .
FT VARIAT 1349    P -> T (IN STRAIN 129/SVJ) .
FT VARIAT 1352    EG -> OR (IN STRAIN 129/SVJ) .
FT VARIAT 1381    S -> P (IN STRAIN 129/SVJ) .
FT VARIAT 1390    G -> A (IN STRAIN 129/SVJ) .
FT VARIAT 1400    V -> D (IN STRAIN 129/SVJ) .
FT VARIAT 1503    E -> Q (IN STRAIN 129/SVJ) .
FT VARIAT 1549    V -> A (IN STRAIN 129/SVJ) .
FT VARIAT 1680    T -> K (IN STRAIN 129/SVJ) .
FT VARIAT 1712    D -> E (IN STRAIN 129/SVJ) .
FT VARIAT 1721    D -> E (IN STRAIN 129/SVJ) .
FT SEQUENCE 1812 AA; 198669 MW; 2291EA74150BB86A CRC64;

Query Match      49.9%; Score 4817; DB 1; Length 1812;
Best Local Similarity 55.5%; Pred. No. 2.9e-194;
Matches 1035; Conservative 261; Mismatches 498; Indels 70; Gaps 32;

QY 1 MDLSALRVEEVONVINAMQKILECPICILELIKEPVSTKCDHIFKFCMLKLLNKKGPSP 60
DB 1 MDLSAVQIOEVQNVLHAMQKILECPICILELIKEPVSTKCDHIFKFCMLKLLNKKGPSP 60

QY 61 CPLCKNDITKRSLOSTRFVSOLVRELLIIICAFOLDTGLEVANSYNPAKKNNSPEHLKD 120
DB 61 CPLCKNEITKRSLOSTRFVSOLVRELLIIICAFOLDTGLEVANSYNPAKKNNSPEHLKD 120

QY 121 EYVSIQSGYRNRAKRLQSGEPNPISLOETSLVSQSLNGLTVRTLRKTQRIQPKTSYVI 180
DB 121 EASIIQSGYRNRRRLPQVEPNATKLD-SLGVQSLNGLIVRSVKKNNROTQPRKKSYYI 179

QY 181 ELGSDSSEDTYNKATYCSVGQELLOITPQGTREISLDSAKKACFESETDVTNTEHQ 240
DB 180 ELDSDSSEETVTKPGDCSVRQOELLQTAQAGDEGLHSAEAEACFESE-GIRNIEHQ 238

QY 241 PSNNDLNTEKRAARHPKPYQGGSVSNLHVPCGTNTNTHASSLOHENSLLLTCKDRMVE 300
DB 239 CS-DOLNTPENHATERHEPKCQSTISINVCVPCGTDAHASSLOPETSSLLLIEDRNAE 297

QY 301 KAEFCNKQKGLARSQHRNAGSKETCNDRTPTSTKQVLDNADPLCEKRWKQKLPC 360
DB 298 KAEFCNKTKQGVIAVSQSRWPAASKGTCDNRQVPSTGEKVGPNADSLSDREKWHQPSLC 357

QY 361 SENPROTEDEVNITLNSSIOKWNHFSRDELLGSDSDSHCESNAKAVADVLVNEVD 420
DB 358 PENSGATTDVPMITLNSSVQKWNHFSRDELLGSDSDSHCESNAKAVADVLVNEVD 417

QY 421 EYSGSSEKIDLLASDPHEALICKSRVRSKSVESNIEDKIFGKYRKKASLPNLSHYTEN 480
DB 418 GGFSSSRKTDLVTPDHTLMLCKSRDPSKPYEDNISDKIFGKYQRKGRPHLNHWTE- 476

QY 481 LIIGAFVTEPQIOBRPLTNKLRKRRTPSGLHPDEFIKKADLA-VOKTPMIGNQTNQT 539
DB 477 -IIGTFITEPQIOBRPLTNKLRKR--STSLQPEDFIKKADSAGVQRTPDNINQGTDL 533

QY 540 EQNGQVMNTNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSISINMWELE 599
DB 534 EPNEQAVSTTSCNQENKTAGSLNLOKESAHPTESLRKEPASTAGAKSISINVSVDLEVL 593
```

```
QY 600 IHNSKAPKKNRLRRKSSSTRHIALELVVSRNLSPNCTELQIDSCSSSEIEKKKKYNQMP 659
DB 594 VHSSKAPKKNRLRRKSSIRCALPLE-PISRNPSPTCAELQIDSCSSSEIEKKKNISNOQA 652

QY 660 VRHSRNLQLMGKEPATGAKSKNKPNEQTSKRHDSDTFPELKLTNAPGFTKCSNTSELK 719
DB 653 AGHLREPQLIETEDTEPAADAKK-NEPNEHIIKRRASDAFFPEEKLNMKAGLLTSCSPRKSQ 711

QY 720 EFNVPSPREEKEKLETYKVSNNAEADPKDMLSGE-RVLOTERSVSSSISLVPCTDYG 778
DB 712 GPVNPS-PQRTGTQLETRQMSDAKELGDRVLGEGPSGKTTDRSEESTSVSLVSDTDYD 770

QY 779 TOESISLEVSTLCKATEPNKVCQCAAFENKGLHGCCKDNRDTEGFKYPLGHEVN 838
DB 771 TQNSVSVLDHAHTVRYARTGSAQCMQFVASENPKELVHG-SNNAGSGTEGLKPLPRLHALN 829

QY 839 HSRETSIEMEESELDAQYLONTFKVSKRQSFALFSNPGNAEEECATFSAHSGSLKQSPK 898
DB 830 LSQE-KVEMEDSELDQYLONTFQVSKRQSFALFSKPRSPQKDC---AHSVPKSLSPK 884

QY 899 VTPECKEENOGKNEINIKPVQTVNITAGPPVVGQDKPVDNAKCSIKGSRFCLSSQF 958
DB 885 VTAKGQK-ERQGOEFEIHSVQVAATAVGLVPCQEGKLAADTMC--RGCRLCPSSH 941

QY 959 RGNETGLITPNKHGLLQNPYRIPPLPIKSVKTKCKKNLLEENFEHSMSPEREMGNEN 1018
DB 942 RSGENGLSATCKSGIISONSHFKQSVPIRSIKITDNKPLTEGREFERTSSTEMAVGNEN 1001

QY 1019 I-PTVTTISRNNTIRENVFKEASSNINEVGSSNEVGSSINEIGSSDENTQOALGNRG 1077
DB 1002 ILQSTVHTVSLNN-RGNACQEAGS-----GSIHEVCSTGDSFPQGLGNRG 1046

QY 1078 PKLNAMRLGVLOPEVYKQSLPGSNCKHPKIKQBEYEVQTVNTDPSPLYLISDNLEQPM 1137
DB 1047 PKVNTVPLDSMQGVQCQSVPVSD-KYLETKQGEAVC---ADFSPLCFSDHLEQSM 1101

QY 1138 GSSHASQVCTPDLLDODGEIKEDTFAENDIKESAVFSKSVQKGLSRLSPSPFTTH 1197
DB 1102 -SGKVQVCSTPDLLDDVEKQHTSFGEDIMERSAVFNGLSRLRESSPSVPVTHAS 1160

QY 1198 LAQYRGAKKLESSENLSEDEELPCFQHLFGKVNINIPQSTRHSTVATECLSKNTE 1257
DB 1161 KSSLHRSARSKLESSESDSTEDDLPCFQHL--SRISNTP-ELTRCSSAVTQRMPEKAE 1218

QY 1258 ENLLSLKNSLNDCSNOVILAKASOEHLHSETKCSASLFSQCSLELDLTANTQDPL 1317
DB 1219 GTQAPWKGGSSDCNVEIMIEASQEHQFSEDPKCSGSMFSSQHSQSAVOGSTPNANSQDNF 1278

QY 1318 IGSSKQMRHQSESQGVGLSDKELVSDDEERGTLGEENNOERQSDMSNLGEA-ASGESET 1376
DB 1279 IPPSTQRSHQCGNEEAFLSDKELISDNEMATCLEEDNDQEE--DSIIPDSEASGESET 1336

QY 1377 SVSEDCSGLSQSDILTTQORDTNQHNLIKLOQEMAELEAVLEQHGQSPNSYPSIISDS 1436
DB 1337 TLSDEC---SQSDILTPQEGATMKYNLIKLOQEMAELEAVLEQHGQSGSHPSLLGDP 1392

QY 1437 SALEDLRNPEQSTSEKAVLTQSKSSEYPISONPEGLSADKFEVS-ADSSSTKNKEPGYER 1495
DB 1393 CALEDLPVLEPNMGGAAILTSKNINENPVSONLKASACDKPQLQHLGPTSGDDESGMR 1452

QY 1496 SSPSKCPLSDDRWTHSCSGLQNRNYPQBELIKVYDVDEBQOLESOPHDLTETSUPR 1555
DB 1453 PSPFKPLAGSRGSAHCSRHLQKRNPSQOELLQPAGE---ASSEPHNSTGESCLPR 1508

QY 1556 QDLGTPYLESGISLPSD-DPESDPSEDRAPESARVNIPSTSAKVPQLKVAESAQSP 1614
DB 1509 RELEGTPLGSGISLFSRDPPESESPK---EPAHIGTTPASTSVLKTPQOCQAFRSAA 1564

QY 1615 AAATTTDTACYNAMEEVSREKPELTASTERVNKRMSVYSGLTPEEFMLVYKFAKHHI 1674
DB 1565 AGAD-----KAVGVIGVSKIKPELTSEERADRDIMNVVSGLTPEKVMVQKFAEKYRL 1617
```

QY 1675 TLNLITEETHVVMKTDAEFVCERTLKYFLGIAGGKVVYFWVTQSIKERRMLNEHDF 1734
Db 1618 TLTAITEETHVVIKTDAEFVCERTLKYFLGIAGGKVVYFWVTQSIKERRMLNVEH 1677
QY 1735 EVRGDVNNGRHQPKPARESQRKIFRGLIEICCYGFTNMPDQLEWVQLCGASVYKE 1794
Db 1678 EVGTGVVTVGRNHQPPRRSRE-KLFKGLQVYCCDPFTNMPKDDLERMLQLCGASVYKE 1736
QY 1795 LSFTLTGTGHPVIVVOPDAWEDNGHFAIGQCEAPVTVREXWVLDVALYQCELDLYL 1854
Db 1737 LPSLTHTDGAHLVIVVQPSAWEDNSNCPDGLQJCKARLVMDWVLDSSLRYCRDLDAYL 1796
QY 1855 IPQI 1858
Db 1797 VQNI 1800

RESULT 4
BRC2_MOUSE STANDARD; PRT; 3329 AA.
AC P97929; P97383; O35922;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
GN Breast cancer type 2 susceptibility protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Connor E., Smith A., Wooster R., Stratton M., Dixon A., Campbell E.,
RA Trait T.M., Freeman T., Ashworth A.;
RT "Cloning, chromosomal mapping and expression pattern of the mouse
RT Brc2 gene.";
RN Hum. Mol. Genet. 6:291-300(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=97237041; PubMed=9119389;
RA Sharon S.K., Bradley A.;
RT "Murine Brc2: sequence, map position, and expression pattern.";
RN Genomics 40:234-241(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=97384941; PubMed=9242436;
RA McAllister K.A., Haugen-Strano A., Hagevik S., Brownlee H.A.,
RA Collins N.K., Futreal P.A., Bennett L.M., Wiseman R.W.;
RT "Characterization of the rat and mouse homologues of the BRC2 breast
RT cancer susceptibility gene.";
RN Cancer Res. 57:3121-3125(1997).
RN [4]
RP SEQUENCE OF 18-200 FROM N.A.
RX MEDLINE=97075121; PubMed=8917547;
RA Rajan J.V., Wang M., Marquis S.T., Chodosh L.A.;
RT "Brc2 is coordinately regulated with Ccr4 during proliferation and
RT differentiation in mammary epithelial cells.";
RN Proc. Natl. Acad. Sci. U.S.A. 93:13078-13083(1996).
RN [5]
RP SEQUENCE OF 569-625 FROM N.A.
RX MEDLINE=97341126; PubMed=9196008;
RA McAllister K.A., Ramachandran S., Haugen-Strano A., Fiedorek F.T. Jr.,
RA Wiseman R.W.;
RT "Genetic mapping of the Brc2 breast cancer susceptibility gene on
RT mouse chromosome 5.";
RN Mamm. Genome 8:540-541(1997).
CC -!- FUNCTION: MAY PARTICIPATE IN A PATHWAY ASSOCIATED WITH THE
CC ACTIVATION OF DOUBLE-STRAND BREAK REPAIR AND/OR HOMOLOGOUS
CC RECOMBINATION (BY SIMILARITY).

CC -!- SUBUNIT: INTERACTS WITH RAD51 (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST EXPRESSION IN
CC CEREBELLUM, TESTIS, ILEUM, APPENDIX, EPIDYMIIS, OVARY AND MAMMARY
CC GLAND. NO EXPRESSION IN LUNG.
CC -!- DEVELOPMENTAL STAGE: IN THE MAMMARY GLAND, EXPRESSION INCREASES
CC DRAMATICALLY DURING PREGNANCY.
CC -!- SIMILARITY: CONTAINS 7 BRC2 REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U82270; AAB48306.1; -;
DR EMBL; U72947; AAB40720.1; -;
DR EMBL; U65594; AAC23702.1; -;
DR EMBL; U89652; AAB71377.1; -;
DR EMBL; U89503; AAC53276.1; -;
DR MGD; MGI:109337; Brc2.
DR InterPro: IPR002093; BRC2_repeat.
DR Pfam: PF00634; BRC2; 7.
DR PROSITE: PS0138; BRC2_REPEAT; 6.
KW Polymorphism; Repeat.
FT REPEAT 981 1015 BRC2 1.
FT REPEAT 1192 1226 BRC2 2.
FT REPEAT 1394 1428 BRC2 3.
FT REPEAT 1491 1525 BRC2 4.
FT REPEAT 1623 1657 BRC2 5.
FT REPEAT 1924 1958 BRC2 6.
FT REPEAT 2004 2038 BRC2 7.
FT VARIANT 44 44
FT VARIANT 340 340
FT VARIANT 377 377
FT VARIANT 407 407
FT VARIANT 661 661
FT VARIANT 739 739
FT VARIANT 1038 1038
FT VARIANT 1198 1198
FT VARIANT 1257 1257
FT VARIANT 1392 1392
FT VARIANT 1520 1521
FT VARIANT 1583 1583
FT VARIANT 1613 1613
FT VARIANT 1686 1686
FT VARIANT 1799 1799
FT VARIANT 1881 1881
FT VARIANT 1894 1894
FT VARIANT 2141 2141
FT VARIANT 2392 2392
FT VARIANT 2605 2605
FT VARIANT 2648 2648
FT VARIANT 2717 2717
FT VARIANT 2729 2729
FT VARIANT 2814 2814
FT VARIANT 2827 2827
FT VARIANT 2907 2907
FT VARIANT 2929 2929
FT VARIANT 3058 3058
FT VARIANT 3071 3071
FT VARIANT 3081 3081
FT VARIANT 3089 3089
FT VARIANT 3105 3109
FT VARIANT 3220 3220
FT VARIANT 3238 3238
FT VARIANT 3243 3243
FT VARIANT 3245 3245
SQ SEQUENCE 3329 AA; 370658 MW; A3B269AF8F05A5D1 CRC64;
Query Match 3.7%; Score 357.5; DB 1; Length 3329;
Best Local Similarity 19.2%; Pred. No. 2.2e-07;

Matches 401; Conservative 292; Mismatches 742; Indels 649; Gaps 94;

QY 72 SLOESTRESQLVBEELKLTIC-----AFOLDTGLYANSYNFAKKENNS---PEHLKD 120
 Db 466 SLEDSTAGQWVSRTSQAACLSPTSRKSLFKWREPLDETGLTVFSDSMTNFTTEHREAS 525
 QY 121 EVSIIOSMGYRNRKAKRLQASEPNFS---LOETSLSVQLSGLGTVRTLRKQRIQPOKT 176
 Db 526 ACGLGILTACSQREDISCFSSVDTSWPTTLTDSATVK--NAGLISTLKNKR----- 577
 QY 177 SVYIELGSDSSEDTVKNATYCSVGDELLQIIPQGRTRBISLDSAKKACBSETDV--- 233
 Db 578 -----KFTY--SVSDASIQGKKLQTHROLELTNL--SAQLEASAFEVPLT 619
 QY 234 -TNTHEHQPNSNDLNTTEKRAAERHEPKYQSGSVSNLHVPECGTNTHASSLQHENSSL-- 290
 Db 620 FTWVNSGIPDSD---KRCPLNDPE-----EPLSTNSFGTATSKELSYIHA 663
 QY 291 LLTKDRMNVKAEFCNKSQOPGLARSQHNRWAGSKETC--NDRRTPEST---EKKVDLNA-- 344
 Db 664 LISQD-LQDKEATVIEBKPPVTAREADFLCLPERTCENDQKSPKVSNGREKVLVSACL 722
 QY 345 -----DPLCKERWKKQL--PCSENPROTEDEVW----- 372
 Db 723 PSVAQLSSISFESQENPLGDHNGTSLKLTPSSKLPFLSKADMVSRKMKMPEKLQCESC 782
 QY 373 ---ITLNSSIOKVNEWFSRSDLLGSDSDHGESSENAKV-----ADVLDVLNEVDEY 422
 Db 783 KVNIELSKNILEINEICILUSE-----NSKTPGLPPGGENIIEVASMSKQ 827
 QY 423 SGSEKIDLLASPDHEALICKSER-----VHSKSVESNIEDKI-----FGKYRKAKSLP 472
 Db 828 FNQNAKI-----VLQDKQSPPISEVAVNVNNEELFPDGSNNFAQVYNKCNKP 877
 QY 473 NLSHTVENLIIAGFVTEPOIIORPL--TNKLUKRRRTPSGLHPEDFTKKADLAVOKTPE 530
 Db 878 DLGSSVE-----LOEEDLSHTQGPSPKSNPMVADEDDVDDAHAAQVLIITKDS 924
 QY 531 MINOGNTOEQGVNMNITNSGHNKT--KGDSIONEKNPNPTLESLEKESAFKTRAEPTSS 589
 Db 925 SLAVVHDYTEKSRNIEHQKOTEDKDFKSNSSLNKSDGNSDCSDKWSEF---LDPVLN 981
 QY 590 -----SISNMLELNTNHSKAPKK--NRLRRKSSTR-----HIHALELVVSRNLSPP 634
 Db 982 HNPFGSFRASNKEIKLSEHNKVKSMFFKIDEEQYPTRLACIDIVNTLPLANQKKISEP 1041
 QY 635 NCTELQIDCSSSEELKKKKYQMPVRHSRNLQLMGEPATGAKSNKPNEQTSKRHDS 694
 Db 1042 HIFDLK-----SVTTVSTOSHNOSSVSH-----EDT 1067
 QY 695 DTPPELKLTPNAPGSFTKCSNTSELKEFVNPSUPREKEEKLFTVK-----VSNNAEDP 747
 Db 1068 DTAPQM--LSKQDFHNNLTTSQKAETTELSTILEESSQOFTQFRKPSHIAQNTSEVP 1126
 QY 748 KDMLSGERVLOT--ERSVRESSISLVPQTDYGTQESISLLEVSTLGLKAKTEPNKCVSOCA 806
 Db 1127 GNQWV---VLSTASKWKDLDLHPDVPSPVQTDHSHQFEGSA--GVKQSPFHLLLEDTCN 1181
 QY 807 APENPKLLHGCSKDNRNDE--GFKYPLGHEVNHRSRTSITEMESELDAQYLONTFKVS 864
 Db 1182 --KNT-----SCFLPNINMEFGGFCALGTKLVSNS-----A 1213
 QY 865 KROSFAFNSPNAEECA-----TFSAHSGS-----LKKQSPKVTFE-----CEQK 906
 Db 1214 LRKAMKLFSDIENSEEPSAKVPGRFSSSAHHDSVASVFKTKKQNTKESKDFEKSCKQV 1273
 QY 907 EENQ-----GKN--ESNIKPVQ-----TNVITAGPPVW---GQK 935
 Db 1274 LQNNIEMTTCIFVGRNPEKYIKNTKHEDSYTSSQNNLENSDGSMSTSG--PVYIHRGDS 1332
 QY 936 DRPVD--NAKCSKSGSRCLSQFRNGTGLITPNKHGLLQNPRIPLPIKISFVKTKC 994
 Db 1333 DLPADQGSKC-----PESC--TOYARENTQIKENISDLTCLTEIMKABETCMKSSDKQL 1385

QY 995 KKNLLEENFEHSHMSPEREMGNIPSTVSTISRNNIRENVEKFEASSSNINEVGSSTNEV 1054
 Db 1386 PSDKMEQNIKENFIS-----FQTASGNTR--VSKESLNKSVNFINRETDE- 1429
 QY 1055 GSSINEIGSSDENITQAEELGRNRPKLNAMLRLGLVLOPEYIKOSLPGSCNKHPEIK--KQBY 1113
 Db 1430 ---LTVISDS-----LNSKILHGINKDKMH-----TSCHKKAISIKKVF 1465
 QY 1114 BEVQTVNMTDPS-----YLISDNLEOPMGSSHAS-----QVCSETPD---DLLDDGE- 1158
 Db 1466 EDHFPITVTSQALPAOHPEYIEISTKEPTLLSFHTFASGKKVKIMQESLQKVNLFDETOY 1525
 QY 1159 IKEDTSPAEND--IKESSAVFSKSVQKSGELSRSPFTHTLAQGYRRKAKKLESEENL 1216
 Db 1526 VRKTASFSGQSKPLKDSKKELTLAYEKTEVT-----ASKCEMONEV 1567
 QY 1217 SSEDELPFCFOHLLFGKYNVNIQSQSTRHSTVATE-----CLSKN---TEENL 1260
 Db 1568 SKETEMLPQONHYMYRQTEENLKTNGTSKVOENIENNVEKNPRICCIQCSSYPVTEDSA 1627
 QY 1261 LS--LKNSLNDCSNOVILAK-----LSEETKCSASLFSFSSOCSELDL----- 1306
 Db 1628 LAYTETDSRKTCVRESSLSKGRKWLREOGDKLGRNTIKIECVKHEHTDFAGNASYEHS 1687
 QY 1285 -----LSEETKCSASLFSFSSOCSELDL----- 1306
 Db 1688 VIIRTEIDTNHVENQVSTLLSDPNVCHSYLSQSFCHCDMDHNDSGYFLKNNKIDSDVPP 1747
 QY 1307 -----TAN-----TNTQ-----DPLIGSSK-- 1322
 Db 1748 DMKNAEGNTISPRVSATKERNLHPOTINEYCVQKLETNTSPHANKDVAIDPSLLDSRNCK 1807
 QY 1323 -----QMRHOSQSGVGLSDKELVSDDEERGCTGLEENNQBEQSMDSNLGEAASCEST 1376
 Db 1808 VGSVLVFIHAHQETE---RTKEIVTDNCYKIV-----EQNRQSPDTQTSCHKVL 1855
 QY 1377 SVSBDSCSLQSOSDILLTTOQDRT-----MOHNLKIQQEMAELEAVLEOHGQSOPS 1428
 Db 1856 DDKDFCIPSSSGDVCINSRKDSFCPHNEQILOHN-----QSMS-----GLKKAAT 1901
 QY 1429 YPSIISDSSALEDLRNPEQ-----STSKAVLTQSKSSEYPTSQNPGLSADK 1476
 Db 1902 PPVGLTWDTSKSIREFPPQAAHPSRTYGFSTAGKAIQVSDASLE--KARQVFESEMDGDA 1960
 QY 1477 FEVSADSTSKNKEP--GVERSSSPKCPSLDWRVYMHSCGSLQ-----NRNYP---- 1523
 Db 1961 KQLSSMVSLGNEKPHHSVKRENS-----VVHSTQGVLSLPKPLPGHNVNSVFSG 2010
 QY 1524 -----SQBELIKVDV-----EEQOLEESG--PHDLTETSYLPRQDLEGTPY 1563
 Db 2011 FSTAGGKLVTVSESAHLHKVKGMLHEFDLIRTEHTLQHSPIEDVSKILPQPCAEIRTPPY 2070
 QY 1564 -LESIGISLFDSDPSEDRAPEARSVGNIPSPSTSAALKVPQLKVAESAQAASPAHAHTDT 1622
 Db 2071 PVNSKLQKTYNDKSSLSLPSNYK--ESGSGNTQSTIEVLSQLSOMERNQDQTLVLGTVKSHS 2128
 QY 1623 AG-----YNAMESVSREKPEL--TASTERVKNRKMVVSGLTPPEEFMLVYKFAKHHTL 1676
 Db 2129 KANLLGKEQTLQPNIKVKTDEMTFTSDVPKTNVGEYSKESENYFETEAVESAKAFMED 2188
 QY 1677 TNLITEETHVMKTDAEFCVCELTLYFLGIAGGKWWVSFWVTQSIKERKMLNEHD--F 1734
 Db 2189 DELTDSQTHAKCSL---FTCPQNETLNSRTRKGGVVDVAVGQPPIKRSLNEFDRII 2245
 QY 1735 EVRGDVVNGRNGHOGPKRARES---QDRKIF---RGLEICCYGPF 1772
 Db 2246 ESKCKSLT-----PSKSTPDGTVKDRSLFTTHMSLEPVTGCPF 2283

RESULT 5
 RBPL_PLAYB
 ID RBPL_PLAYB STANDARD: PRT: 2869 AA.

Db 1875 EMMKVSAYEGMKRDH7VSQVODMKTIIIVDELKTLNDISECSVLNVVSVIVKKVKES 1934
 QY 1284 HLSEETKCSASLFSQCSLEDLTANTWTPDPLIGSSQMRHOSQO-----GV 1333
 Db 1935 KHADYRRDANSWYESWVI-----LANVFLSDEAKISSGMFNAEMKSNFKTDLELFSV 1989
 QY 1334 GLSDKELVDDPERGTGLEENNQQSDMSNLGPAAGCSETSVSEDCSLGQSQSDILT 1393
 Db 1990 ISNSNELLKKIEQDSNDVIQERESQELAKDATDIYVNIKLFNEFKLEBAKKEEVS 2049
 QY 1394 TOORTMQ-----HNLKLOQMAELEAVLBOHQSQPS----- 1426
 Db 2050 EKVRALKRLSQVEGIRCFHFRLHLDLELENLKKWTVIYRDKKSERESGLQEMENE 2109
 QY 1427 -NSYPS-----IISDSSALEDLRNPQST-----SEKAVLTOKSSP-----Y 1463
 Db 2110 MNTYNSITQLEGIVVSGESKEIEKLNERNEMRNISEKISTIDSKVTMNSTIDELY 2169
 QY 1464 PISQNPGLSADKFEVSADSTSK-----NKEPGVERSPSK-----PSLDDRWTMHS 1512
 Db 2170 KLGNKQAHWISLISYANTMKTSSKLMINK-----EKENTKCVDIKDNSSSDTGIVET 2225
 QY 1513 CSGSLQNR-NYPSQELKVVVV-----EEQLESQPHDLTETSYLPRQD-----L 1558
 Db 2226 LKGFYGSKLTFSSAIEYQONADTYSVNFAPKHEKESL--NAIRDIKKELYLFHQNSDISIV 2283
 QY 1559 ECTPYLESIGILFSDDPSPEDRAPESAR-VGNIPSTTSALKVPOLKVAESAQSPAAA 1617
 Db 2284 EGG--VQNNLALY-----DKLNEEKREMDELYRNTSETKLQNE----- 2320
 QY 1618 HPTDTAG-----YNAMESVSREKPEITASTERYNKR-----SMVVSGL--TPEEPMV 1665
 Db 2321 HSTDVFKPMIELHKGMMNETNKKSLLEKEKLVSDNHMSMEAEWIKNLKTYTPESVONI 2380
 QY 1666 -----YKFAKHH-----ITLTN-----LITEETHVV 1688
 Db 2381 NNIVSVIAEAVKTLFEIDRDYGDYQIVVEEHKQFSILIDRTNALMDDIEIFKKENNYL 2440
 QY 1689 MKTDAEFYCERLKYFLGIAGKVV--VSFYWVTQSTIKERKMLNEHDFEVRGVDVNGRN 1745
 Db 2441 MEVNTETI-HRVNDYIEKIT-NKLVAQTEYEQILENIKQNDMLQNIPLFKKYSII--EY 2496
 QY 1746 HOGPKRARES 1755
 Db 2497 FENVKVKES 2506

RESULT 6

YD86_SCHPO STANDARD: PRT: 1957 AA.
 AC Q10411;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein c1f3.06c in chromosome I.
 GN SPAC1f3.06c.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=118593560;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 270690; CAA94624.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;
 Query Match 3.5%; Score 339; DB 1; Length 1957;
 Best Local Similarity 18.5%; Pred. No. 6.7e-07;
 Matches 354; Conservative 325; Mismatches 650; Indels 548; Gaps 79;
 QY 3 LSALRVEEVONVINA-----MQTLECP-----CLELIKEPVSTKCDHIFCKFCMLKL 51
 Db 266 VSTLR--QTENSLRAECKTLQEKLEKCAINEEDSKLLEELKHNVANYSDAIVHKDKLIED 323
 QY 52 LNOK-----KGPSQCLPKNDITKR-----SLOES-TRFSOLVEELKLIICAFOLD 96
 Db 324 LSTRISFEDNLKSERDPLSLKNEKLEKLLRNTIGSLKDSRTSLSOLEEEMVEL----- 376
 QY 97 TGLEVANSYNFAKKENNSPEH--LKDEVSITQSMYGNRAKRLQLQSEPNPSLOETSLV 154
 Db 377 -----KESNRTIHSQTLDAESKLSFEQENKSLKSGIDEYQNNLSSKDKVVK 423
 QY 155 QLSN-----LGTVRTLTQRIOQKTSVYIELGSD-----SSEDVTNKA 195
 Db 424 QVSSQLEEARSSLAHATGKLAENISERDFQNKIKDFEKIQDLRACLNSSSNELKEKSA 483
 QY 196 YCSVGDOELLQITPOGTRDEISLSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAE 255
 Db 484 LIDKQOELNNLRBQ--IKEQKKVSESTQSSLSQSLQORDILN-----EKK--- 525
 QY 256 RHPEKYOGSSVSNLHVPCGTNTTHASSLQHENSSLLTTKDRMNVKAEFCNKSQOPGLAR 315
 Db 526 KH-EVYE-SQNLKGE---LQTEISNSELSSOL----- 555
 QY 316 SQHNWAGSKETCNDRTPTSTKVKVDLNADPLCE--RKENWKQLPCSENPRDTEVPWI 373
 Db 556 ---STLAAEKEAAVATNNELSEKNSLQT--LCNAPQELAKSVMLKENQDN-----FS 605
 QY 374 TLNASSQKVNWFERSRDELGSDSHGSESNKAVADVLDVLN-----EVDEYSGSSE 427
 Db 606 SLDTSFKKLNE---SHOEL---ENNHTIYKQLKDTSSKLOQLQELRANFEQKSTLSDE 659
 QY 428 KIDLLA-----SDPHEALICKSERVSHKSVESNIEDIFGKTYRK-----KASLNL 474
 Db 660 NNDLRYKLLKLEESNKSLIKQEDV--DSLEKNIQ--TLKEDLRKSEALRFSKLEAKNL 715
 QY 475 SHVTENLIIGAFTVEPTQIIQERPLTNKLRKRRTPTSGLHPEDFTKKADLAVQKTP--EMI 532

Db 716 REVINDL---KGGHETLEAQRNDLHSLSDAKN-TNAILSELTKSSDEDVKRLTANVETL 771
QY 533 NQGTNOTEQN-----GOVMNITNSCHENKTKGDSIQNEKNPNPTIESLEKESAFTKAPEI 587
Db 772 TQDSKAMQSFSLVNSQISLNLHELRODHVNMQSONN-----TLLESSEKLTDCENL 827
QY 588 SSS-----ISNWEELNTHNSKAPKKNLRKRSKSPRHHALELVYSRNLS----- 632
Db 828 TQQNMTLIDNVQKLHMKHVNQESKVSLEKEVNGKLSLDLKNLRSNLVAISDNDQIILTQL 887
QY 633 ---PPNCTELQIDSSSEELKKKKYNOMPVRHRSRNQL-----MEGKEPATCAKSKNK 683
Db 888 AELSKNTVDSLEQESAQLNSGLSLAEKQ-LLHTENELHRLDKLTKGKIEESKSDL 946
QY 684 PNEOTSRHRSDTPELKLTAAPGFTKCSNTSELKFEVNPSPPREKEE-----KLSTVK 739
Db 947 GKLTARQEE---ISNLKEENMSQAITSVKSKLDETLSKSKLEADIEHLKKNVSEVE 1003
QY 740 VSNNAEDPKMLSGERVLOQTERSVSSISLVPDGTQGTQESISLLEVSTPLGKAKTEPN 799
Db 1004 VERNA-----LLASNERLMD----- 1018
QY 800 KCVSOCAAFENPKGLIHGCSKDNNDTEGPKYPLGHEVNHSHRETSIEMESSELDAYQLON 859
Db 1019 -----DUKNNGENTA-----SLQTEIEKKRAENDD--LQS 1046
QY 860 TKFY--SKRQSFAFSLNPGNAEECATFSAHSGSLKKQSPKVTPECKQKBNQK--NES 915
Db 1047 KLSVVSSEYENALLISSQTN-----KSLDKTNQLKYIEKNVQKLDEK 1090
QY 916 NIKPQVTNITAGPPVVGQDKPKVDNAKCSIKGGRSCLSSQFRGNETGTITPNKHGLLQ 975
Db 1091 DORNVEELTSKYKGLGEENAQIKDELLALRKSK-----KQHDICA 1133
QY 976 NPYRIPPLFPKIPKVKTKCKKNLLEENFEHSHSPEREMGNENIPSTVTSIRNNIRENV 1035
Db 1134 N-----FV-----DOLKEKSDALE-QLTNEKNELLVSLQSSNNNEAL 1170
QY 1036 FKEAS--SSNINEVGSSTNEVSGSINEIGSSDENIQELGRNRRGPKLNAMRLGLVQPEV 1093
Db 1171 VEERSDLANLSDMKKSLSDSNVISRSDLVVRNDEL-----DT 1211
QY 1094 YKQSLPSGCKHPEIKKQEEVVEVOTV---NTDFSPYILSDNLEBQPMGSSHASQVCSETP 1150
Db 1212 LKKDKDSLSTQYSEV-CQDRDDLSDLSKGCEESFNKYAVS-----LRELCTPKE 1259
QY 1151 DD-----LLDDGETKEDTSPA-----ENDIKESSAV-PSKSVQKGELSPSPFT 1194
Db 1260 IDVPVSEILDNDNFVNAGNFSELSRLTVLSLENYLDFAFNQVNFKKMELDNRLFTTDAEFT 1319
QY 1195 -----HTHLAQGYRRG--AKKLESSEENLSSEDEELPCFQHLFGKVNPNIPSQST 1242
Db 1320 KVVADLEKLOHEDDHLIQRGDEKALKDSEKNFLRKEAEM-----TENIHSLEE 1369
QY 1243 RHTVATEC--LKNTEENLLS---LKNSLDNCNSQVILAKASEHLSBETKCSASLFS 1297
Db 1370 GKETKEIEAELSRLEDNQLATNKLKNQLD-----HLNQEIRLKEDVLK 1414
QY 1298 SOGS---ELEDLTANTWTPDPLIGSSKOMRHOSESOGVGLSDKELSDDEERGCTGLEEN 1354
Db 1415 EKESLIISLESLSNQKQESSLLDAKNELEH-----MLDDTSRKNSSLMEX 1461
QY 1355 NQE-EQSMDSNLGAEAGSCESETSVSDECSGLSSQSDILATQORDTMQHNLIKLOQEMAE 1413
Db 1462 IESINSSLDKSFELASAVEKLGALQK----LHSESLSLMENIKSQLQEAKEKIQVDEST 1517
QY 1414 LEAVLEOHGQSPNSY-----PSIISD--SSALELURN--PGOSTSEKAVLTSQKSEY 1463
Db 1518 IQE-LDHEITFASKNNYBEGKLNKDSDIIRLSENIEQNLNLLAEKSAVKRLSTEKESEIL 1576
QY 1464 PISONPGLSADKFEVSADSSSTSKNKEPGVRSPPSKCPSLDORWYMHSCSGSLONRNYP 1523
Db 1577 QFNSRLADLEYHKQSVESELGRSKLK-----LASTTEELQLAENE 1616

QY 1524 SQEBELIKVVDVEEQLEESGPHDLTETSYLPRQDLEGTPLYESGIS----- 1569
Db 1617 RLSLITTRMLDQNO-----VKDLSNKKDSLSDELRLTSLRSEDSVASIQKECKIKSNVTE 1670
QY 1570 -----LFSDDPESDPSDRAPESA-----RVGNIPSTSTSAKVPQLKVAESAQSPAAAHT 1619
Db 1671 SLQDVLTSVQARNAELEDVSRVDKIRRRDRDCEHLSGKLKLLHSQLEEQHETFPRAEQ 1730
QY 1620 TDTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEPLVVKF----- 1668
Db 1731 QRMTQGLFGLKETVKQ-----EKLKKLNLNRQEQQLIPRSSILTVYESYIRDEKEIIV 1782
QY 1669 --ARKKHITITNLITE-----ETTHVMVKTDAEFCERTKLYFLGIAGGKVVVSY 1716
Db 1783 LQERLINGIELSQQLPKGYFGYGFYFKNRVEMLDSFKQQVAKLQFL--AGAEFIVKF 1837
RESULT 7
MLP1_YEAST STANDARD; PRT; 1875 AA.
AC Q02455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MYOSIN-like protein MLP1.
DE MLP1 OR YKR095W OR YKR415.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=93247549; PubMed=8483450;
RT Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RA "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L01992; AAA34783.1; -;
DR EMBL; X73541; CAA51948.1; -;
DR EMBL; Z28320; CAA82174.1; -;
DR PIR; S38173; S38173.
DR SGD; S0001803; MLP1.
KW Coiled coil; DNA repair.
FT DOMAIN 69 487
FT COILED COIL (POTENTIAL).
FT FT DOMAIN 531 1678
FT COILED COIL (POTENTIAL).
FT FT DOMAIN 1834 1866
FT CONFLICT 301 301 R -> A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C906867 CRC64;
Query Match 3.4%; Score 327; DB 1; Length 1875;

RA Nguyen K., Peng Y., Samson C., Schroeder M., Snyder S.C., Steele L.,
RA Stringfellow M., Stroup C., Swedlund B., Swensen J., Teng D.,
RA Thomas A., Tran T., Tran T., Tranchant M., Weaver-Feldhaus J.,
RA Wong A.K.C., Shizuya H., Eyfjord J.E., Cannon-Albright L., Labrie F.,
RA Skolnick M.H., Weber B., Kamb A., Goldar D.E.:
RT "The complete BRCA2 gene and mutations in chromosome 13q-linked
RT kindreds.";
RL Nat. Genet. 12:333-337(1996).
RN [3]
RN SEQUENCE FROM N.A.
RP Hunt S., McMurray A., Williamson H.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RN VARIANT OC HIS-2787, AND VARIANTS ASN-372; MET-1915 AND VAL-2466.
RX MEDLINE-96275740; PubMed=8665505;
RA Takahashi H., Chiu H.-C., Bandera C.A., Behbakht K., Liu P.C.,
RA Couch F.J., Weber B.L., Livolsi V.A., Furusato M., Rebane B.A.,
RA Cardonick A., Benjamin I., Morgan M.A., King S.A., Mikuta J.J.,
RA Rubin S.C., Boyd J.;
RT "Mutations of the BRCA2 gene in ovarian carcinomas.";
RL Cancer Res. 56:2738-2741(1996).
RN [5]
RP VARIANTS ASN-372; ASP-991; SER-1147; MET-1915 AND CYS-2034.
RX MEDLINE-96241588; PubMed=8673091;
RA Couch F.J., Farid L.M., Deshano M.L., Tavtigian S.V., Calzone K.,
RA Campeau L., Peng Y., Bogden B., Chen Q., Neuhausen S.,
RA Shattuck-Eidens D., Godwin A.K., Daly M., Radford D.M., Sedlacek S.,
RA Rommens J., Simard J., Garber J., Metrajer S., Weber B.L.;
RT "BRCA2 germline mutations in male breast cancer cases and breast
RT cancer families.";
RL Nat. Genet. 13:123-125(1996).
RN [6]
RN VARIANT GLO-3095.
RX MEDLINE-96225456; PubMed=8640235;
RA Lancaster J.M., Wooster R., Mangion J., Phelan C.M., Cochran C.,
RA Gumbs C., Seal S., Barfoot R., Collins N., Bignell G., Patel S.,
RA Hamoudi R., Larsson C., Wiseman R.W., Berchuck A., Iglehart J.D.,
RA Marks J.R., Ashworth A., Stratton M.R., Futreal P.A.;
RT "BRCA2 mutations in primary breast and ovarian cancers.";
RL Nat. Genet. 13:238-240(1996).
RN [7]
RN VARIANTS
RX MEDLINE-96225457; PubMed=8640236;
RA Teng D.H.-F., Bogden B., Mitchell J., Baumgard M., Bell R., Berry S.,
RA Davits T., Ha P.C., Kehrer R., Jammulapati S., Chen Q., Offit K.,
RA Skolnick M.H., Tavtigian S.V., Jhanwar S., Swedlund B., Wong A.K.C.,
RA Kamb A.;
RT "Low incidence of BRCA2 mutations in breast carcinoma and other
RT cancers.";
RL Nat. Genet. 13:241-244(1996).
RN [8]
RN VARIANT ASN-2415.
RX MEDLINE-96225458; PubMed=8640237;
RA Miki Y., Katagiri T., Kasumi F., Yoshimoto T., Nakamura Y.;
RT "Mutation analysis in the BRCA2 gene in primary breast cancers.";
RL Nat. Genet. 13:245-247(1996).
RN [9]
RN VARIANT BC ASP-2089, AND VARIANT VAL-3412.
RX MEDLINE-97294396; PubMed=9150152;
RA Vehmanen P., Friedman L.S., Eerola H., Sarantaus L., Pyrhoenen S.,
RA Ponder B.A.J., Mhonen T., Nevanlinna H.;
RT "A low proportion of BRCA2 mutations in Finnish breast cancer
RT families.";
RL Am. J. Hum. Genet. 60:1050-1058(1997).
RN [10]
RN VARIANT BC AND PANCREAS CANCER TRP-554.
RX MEDLINE-98316775; PubMed=9654203;
RA Ganguly T., Dhulipala R., Godmellow L., Ganguly A.;
RT "High throughput fluorescence-based conformation-sensitive gel
RT electrophoresis (F-CSGE) identifies six unique BRCA2 mutations and an
RT overall low incidence of BRCA2 mutations in high-risk BRCA1-negative
RT breast cancer families.";
RL Hum. Genet. 102:549-556(1998).

RN [11]
RP VARIANTS BC L-32; R-53; L-81; R-201; A-211; S-222 AND T-3118.
RX MEDLINE-98272917; PubMed=9609997;
RA Katagiri T., Kasumi F., Yoshimoto M., Nomizu T., Asaishi K., Abe R.,
RA Tsuchiya A., Sugano M., Takai S., Yoneda M., Fukutomi T., Nanba K.,
RA Makita M., Okazaki H., Hirata K., Okazaki M., Fututsuma Y.,
RA Morishita Y., Iino Y., Karino T., Ayabe H., Hara S., Kajiwara T.,
RA Houga S., Shimizu T., Toda M., Yamazaki Y., Uchida T., Kunitomo K.,
RA Sonoo H., Kurebayashi J.-I., Shimotsuna K., Nakamura Y., Miki Y.;
RT "High proportion of missense mutations of the BRCA1 and BRCA2 genes in
RT Japanese breast cancer families.";
RL J. Hum. Genet. 43:42-48(1998).
RN [12]
RN VARIANTS OC PRO-75; HIS-2502 AND HIS-3098.
RX MEDLINE-96255732; PubMed=10486320;
RA Gayther S.A., Russell P., Harrington P., Antoniou A.C., Easton D.F.,
RA Ponder B.A.J.;
RT "The contribution of germline BRCA1 and BRCA2 mutations to familial
RT ovarian cancer: no evidence for other ovarian cancer-susceptibility
RT genes.";
RL Am. J. Hum. Genet. 65:1021-1029(1999).
RN [13]
RP VARIANTS HIS-289; ASN-372; ASP-991 AND VAL-3412.
RX MEDLINE-99254821; PubMed=10323242;
RA Li S.-L., Tseng H.-M., Yang T.-P., Liu C.-H., Teng S.-J.,
RA Huang H.-W., Chen L.-M., Kao H.-W., Chen J.-H., Tseng J.-N., Chen A.,
RA Hou M.-F., Huang T.-J., Chang H.-T., Mok K.-T., Tsai J.-H.;
RT "Molecular characterization of germline mutations in the BRCA1 and
RT BRCA2 genes from breast cancer families in Taiwan.";
RL Hum. Genet. 104:201-204(1999).
RN [14]
RN VARIANTS BC, AND VARIANTS
RX MEDLINE-99138688; PubMed=9971877;
RA Wagner T.M.U., Hirtlenhner K., Shen P., Moeslinger R., Muhr D.,
RA Fleischmann E., Concini H., Doeller W., Heid A., Lang A.H., Mayer P.,
RA Petru E., Ropp E., Langbauer G., Kubista E., Scheiner O.,
RA Underhill P., Mountain J., Stierer M., Zielinski C., Oefner P.;
RT "Global sequence diversity of BRCA2: analysis of 71 breast cancer
RT families and 95 control individuals of worldwide populations.";
RL Hum. Mol. Genet. 8:413-423(1999).
CC -!- FUNCTION: MAY PARTICIPATE IN A PATHWAY ASSOCIATED WITH THE
CC ACTIVATION OF DOUBLE-STRAND BREAK REPAIR AND/OR HOMOLOGOUS
CC RECOMBINATION.
CC -!- SUBUNIT: INTERACTS WITH RAD51.
CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS OF EXPRESSION IN BREAST AND
CC THYMUS, WITH SLIGHTLY LOWER LEVELS IN LUNG, OVARY, AND SPLEEN.
CC -!- DISEASE: BREAST CANCER (BC) IS AN EXTREMELY COMMON MALIGNANCY,
CC AFFECTING ONE IN EIGHT WOMEN DURING THEIR LIFETIME. A POSITIVE
CC FAMILY HISTORY HAS BEEN IDENTIFIED AS MAJOR CONTRIBUTOR TO RISK OF
CC DEVELOPMENT OF THE DISEASE, AND THIS LINK IS STRIKING FOR EARLY-
CC ONSET BREAST CANCER. MUTATIONS IN BRCA2 ARE THOUGHT TO BE
CC RESPONSIBLE FOR SOME INHERITED BREAST CANCER. IT IS LINKED WITH
CC MALE BREAST CANCER.
CC -!- SIMILARITY: CONTAINS 8 BRCA2 REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X95152; CAA64484.1;
CC EMBL; X95153; CAA64484.1; JOINED.
CC EMBL; X95154; CAA64484.1; JOINED.
CC EMBL; X95155; CAA64484.1; JOINED.
CC EMBL; X95156; CAA64484.1; JOINED.
CC EMBL; X95157; CAA64484.1; JOINED.
CC EMBL; X95158; CAA64484.1; JOINED.
CC EMBL; X95159; CAA64484.1; JOINED.
CC EMBL; X95160; CAA64484.1; JOINED.
CC EMBL; X95161; CAA64484.1; JOINED.

DR EMBL; X95162; CAA64484.1; JOINED.
DR EMBL; X95163; CAA64484.1; JOINED.
DR EMBL; X95164; CAA64484.1; JOINED.
DR EMBL; X95165; CAA64484.1; JOINED.
DR EMBL; X95166; CAA64484.1; JOINED.
DR EMBL; X95167; CAA64484.1; JOINED.
DR EMBL; X95168; CAA64484.1; JOINED.
DR EMBL; X95169; CAA64484.1; JOINED.
DR EMBL; X95170; CAA64484.1; JOINED.
DR EMBL; X95171; CAA64484.1; JOINED.
DR EMBL; X95172; CAA64484.1; JOINED.
DR EMBL; X95173; CAA64484.1; JOINED.
DR EMBL; X95174; CAA64484.1; JOINED.
DR EMBL; X95175; CAA64484.1; JOINED.
DR EMBL; X95176; CAA64484.1; JOINED.
DR EMBL; X95177; CAA64484.1; JOINED.
DR EMBL; U43746; AAB07223.1; -.
DR EMBL; 274739; CAA8995.2; -.
DR EMBL; 273359; CAA97728.1; -.
DR Genew; HGNC:1101; BRCA2.
DR MIM; 600185; -.
DR InterPro; IPR002093; BRCA2_repeat.

Query Match 3.4%; Score 327; DB 1; Length 3418;
Best Local Similarity 18.1%; Pred. No. 4.4e-06;
Matches 427; Conservative 330; Mismatches 765; Indels 838; Gaps 107;

QY 2 DLSALRVEEQQVNIAMOKIL-----ECPI-----CLELRPEVSTKCDH----- 41
Db 97 ELDFKLDLGRVNPNSRHSRLRTVTKMDQADDVSCPILLNSCLS--ESPVVLOCTHVTTPQ 154
QY 42 ----TFC-----KF-----H-----CMKLKLNQ 54
Db 155 RKSVCVGSLLFHTPFVKVGRQTPKHSLSGAEDVDPMSWSSSLATPTLSSTVLIVRNE 214
QY 55 KGPSCQCLPKNDIT-----KRSLOSTRFSQVLEELLKXICAFQDITGLEVANS 104
Db 215 EASSETVFP---HDTTANVKSYSFNHDESLKKNDRFIASVTD-----SENTNOREAAS 263
QY 105 YNFAKKENNS-----PEHLKDEVSIIQSNGYRNRAKRLLOSPENPSLOET 150
Db 264 HGFGTSGNSFKVNSCKDHIGKSMNPVLEDEV-----YETVVDTSEED 306
QY 151 SLSSVOLS-----NLGTVRTLTKRIQIOPKTSVYIELGSDSDSEDTVN--KATYCSVGDOE 203
Db 307 SFLCSFKCRKNLQKVTSKTKKI-----PHEANADECESKQVREKYSFVSEVE 359
QY 204 LLOITP-----OQTRDEISLDSAKKAACEF----- 228
Db 360 PNDTDPDLSNVAHQPFESGSGDKISKEVVPSLACENWSOLTLGSLNGAQMEKIPLLIHSSC 419
QY 229 ----SETDVTNTEHHQPSN-----NDLNTTEKRAARHPKRYQG 263
Db 420 QDNISEKOLLDTENKRRKDFLTSNSLPRISLPKSEKPLNEETVVKNRDBEEQHLESHT- 478
QY 264 SSVSNLHVPECPCTHASSLOHENSLL-LFKDRMNVKEAEFCNKSOPGLARSOHRWA 322
Db 479 DCILAVKAISGTSVPASSFOGIIKSPRIRESPKETFNASFSGHMTDPNFKKTEASES 538
QY 323 GSK--ETCNDRR-----TPSTEKKVDLNDAPLCE--RKENWK-- 355
Db 539 GLEINTVCSQKEDSLCPNLIDNGSWPATTTQNSVALKNAGLISLTKKTKNFYIAIHDET 598
QY 356 ----QKLP----- 371
Db 599 SVYGGKKIPKQKSELINCQAOFANAFAPLTFANADSGLLHSSVKRCSQN--DSEEP 655
QY 372 WITLASSITQKNEWFSRSD-----ELLCSD-----DS-----HDG 401
Db 656 TISLTSSEFCTILKCSRNETCSNTNVTISQDLDYKCAKCNKEKQLQFITPPADSLSCLOEQ 715
QY 402 EESNAKVADVLDVLNEY-----DEYSGS-----SEKIDLLASDPHEALICKSR 446
Db 1632 SIFLKVKVHENVEKETAKSPATCYTNQSPYSVIENSALAFYTCSCRKTSVSVSOTSLLEAKK 1691

Db 716 QCENDPKSKVSDIKEEVLAAPVQHSKVEYSDTFQSQSKSLLYDHNASTLILPT- 774
QY 447 VHSKSVESNI-----EDKIFCKTYRKASLNLNLSHTVNTENLIIGAF----- 486
Db 775 --SKDVLNLMISRGKESYKMSDKLKGNNYESDVLTKNIPMEKNQDVCALNENYKNVE 832
QY 487 -----VTEP-----QIIQIERPLTNLKKRRRPTS-----GLHPE-----DFT- 518
Db 833 LLPEKYMVRVASPSRKVQFNQNTNLRVQKNOEETTSISKITVNPDPDSSEELSDNENFVF 892
QY 519 ----KKADLAVQKTP-----MINQTNQTEQNGQVMNTNS----- 551
Db 893 QVANERNNALGNTKELHETDLTCVNEPIFKNSTMVLVYDGTGDKQATQV-SIKKDLVYVL 951
QY 552 GHEHK-----TKGDSIONENPNPDESLEKESAFKTK-----AEPISS-----S 590
Db 952 AEENKNSVKQHIKMTLQDLKSDISLNDIKPEKNNDYMNKNWAGLGLPISNHSFGGSFRT 1011
QY 591 ISNMELELNIIHNSKAPKKNRLRK-----SSTRHIIHALELVVSRNLSPPNC--- 636
Db 1012 ASNKEIKLSEHNI---KSKMFFKDIIEQYPTSACVEIVNTLALDNQKKLSKPOSINTV 1068
QY 637 -----TELQIDSCSSEIEIKKKYNQMPVRRSNL---QLMEGKEPATGAKSKSNPNEQT 688
Db 1069 SAHLQSSVWVSDCKNSHITPQMLFSKQDFNSNHNLTSPQKAEITELSTILESGSQFET 1128
QY 689 SKRDS-----DTP--PELKLTNAPGFTKCSNTSELKEFVN-PSLPRBREKEKLETVKV 740
Db 1129 QFRKPSYILOKSTFEVPEPNQMTILKTTSECRD-ADLHVIMNAPSIGQVDSKOPE--- 1183
QY 741 SNNAEDPKDMLSGERVQLQTERSVESSSISLVPGTQVQTESISLLEVSTLCKKATEPNK 800
Db 1184 -----GTVE-----IKRKFAGLLKNDCK 1202
QY 801 CVSOCAAFENPKGLIHGCKDNMRNDTEGKYPLGHEVNHRSRETSTEMEESLDAQYLQNT 860
Db 1203 SASGLYTDENEVGF-----RGFYSAGHTKLVNSTEA-----LQKA 1237
QY 861 FKVSKRSFALFSPGNAEECATFSHAHSGSLKK---QSPKVTFECE---QKEENQGN 913
Db 1238 VK-----LFSDIENISEETSA-EVHPISLSKSKCHDSVVMFKIENHNKDTVSEKNN 1288
QY 914 ESNIKPQVTNITAGF---PVWGQKDPVDNAKCSIKGSRFCSSQFRG----- 960
Db 1289 KCQLLONNIEMTGTVEEITENYKRNTEENDKYTAASRNSHNLEDFGSDSSKNDTVC 1348
QY 961 ---NETGLIITPNKHGLLQNPYRIPPLPIKSFVK---TKKKNLLEENFEHMSPEREM 1014
Db 1349 IHKDETDLFTDQHNIC-----LKLSGQFMKEGNTQIKEDLSDLTFLEVAKAQEACH 1400
QY 1015 GN-ENIPSTVSTISRNIR-----ENVFKEASSNINEVGSTNEVGSSINEIGSSDENI 1068
Db 1401 GNTSNKQDLTATKTEQNIKDFETSDTFFQTASGNISVAKESFNKI---VNFQDQKPEEL 1457
QY 1069 QAEGLNRGRKPLNAMLGLVQLQPEVYKQSLPGSNCKHPEIKKQE-----YEEVQTVNTD 1123
Db 1458 -----HNSLSNEL-----HSDIRKNKMDILSYEE-----TD 1484
QY 1124 FSPYLISDNLEQPMGSSHASQVCSSETPDDLDDGEIKEDT-----SFAENDIK 1171
Db 1485 IVKHKILKE-SVPVGTGNQLVTFQGPQE---RDEKINEPTLLGFHTASGKKVKIAKESLD 1540
QY 1172 ESSAVFSKSVQKGLSRSPSPFTHLQAQYRGAKKLESSEENLSSEDEELPCFQHLFF 1231
Db 1541 KVKMLFD---EKEOGTSEITSFSH-----QWAKTLKYRE---ACKDLELAC----- 1580
QY 1232 GKVNNIPQSQTRHSTVATECL---SKNTEENLLSLKN-----SLNDC-----SNQ 1273
Db 1581 -----ETIETAPKCKEMQNSLNNKDNLYSIVETVPPKLLSONLCRQTEENLTKTSK 1631
QY 1274 VILAKASOEHLHSEET-KCSASLFSQOC--SELED-----LTANTNTOPDFLIGSSK 1322
Db 1632 SIFLKVKVHENVEKETAKSPATCYTNQSPYSVIENSALAFYTCSCRKTSVSVSOTSLLEAKK 1691

QY 703 TNAPGSFTKCS--NTSELKEPVNPSLPREEKEKLETVKVSNNADPKDMLSGERVLOT 760
Db 2107 KWVP-RFOPTSEHQTRVEQLAN--HLKEKTKDCSELLSKSKE-QLQRDIOERNEEIEKL 2161
QY 761 ERSVSESSISLVPTDGTQBSISLELVSTLGKAKTTPNKVCVSOCAAFENPKGLIHGCSK 820
Db 2162 EFRVRELEQALLVSAD--TFQKVB--DRKHFGAVEAKEPE--LSLEVOQLAERDAIRKEK 2215
QY 821 DNRNDTEGFKYPLGHEVNHRSRETSIEMSELDQAYLQNTFKVKSKROSFALFSPGNAEE 880
Db 2216 EITN-----LEEQLQEPRE-----ELENKNEEVQQLHMQLEIOKKESTRPLQ---ELEQ 2261
QY 881 ECATFS-----AHSGLKKQSPKVTFPE-----CEQKE-ENQCKNESNIKPVQTVN 924
Db 2262 ENKLFKDDMEKGLAIKESDAMSTQDQHVLFQKFAQIIQKEVEIDQLINEQVTKLQOOLK 2321
QY 925 ITAGFPVVGOKDPVDNAKSIKSGRFLCSSLQSPQRNETGLITPNKKGILLQNPRIPLP 984
Db 2322 ITTONKVIIEKNELIRLETOIE-----CLMSD----- 2349
QY 985 PIKSFVTKCKKNLLENFEE-----HMSPER----- 1012
Db 2350 -----QECVKRNEEIEFQNLNEVIEKLOELANICQKTMNAHSUSEADSLKHOLDV 2402
QY 1013 -----EMGNENIPSTVSTISRNIRENVFK-----EASSSNINEVGS 1049
Db 2403 VIAEKLALEQOVETANEEM-----TFMKNVLKETNFKMNQLTOELFSLKRESEVEXIQS 2457
QY 1050 -STNEVGSSINETGSSDENTQAEGRN-----RQPKNAMLRGLVQL 1090
Db 2458 IPENSVNAIDHLSKDKPEPELVLTEDALKSLENQTYFKSFEENGKSGIINLETRLLQLE 2517
QY 1091 PEVYKQSLPGSNCKHPKIKOEVEEVQTVNTDFPYLISDNLPQPMGSSHASQVCSETP 1150
Db 2518 STVSAKOLELTQC--YKQKMDQEQGFETEMLQKKIYNLQKIVEKVAALVSOIQLEA- 2575
QY 1151 DDLDDGEIKEDTSFAENDIKESSAVFSKSYQ-----KGELSRSPSPFT----- 1194
Db 2576 -----VQYAKFCQDNQTISSEPERTNIQNLNQLREDELGSDISALTLRISELESQV 2627
QY 1195 ---HTHQAQYRGCAKLESEENLSEDELPFCFHLLFGKVNINPSQSTRHSTVATEC 1251
Db 2628 VEMHTSLI-----LKEQYVETAEKNVLEKKEKLLQELQLEKNEKKQREKKRSPQDVEV 2693
QY 1252 LSKNTEENLLSKNSLNDSCNOVLAKASOEHHLSEETKCSASLFSQCSLELDLTANTN 1311
Db 2684 LKTTTE---LPHSNEESGFFNEALEARAE-----SVATKAELASYKKAELQLELLVKE 2735
QY 1312 TQDPFLIGSKQMR-HQSES-QGYGLSDKELVSDDER----- 1347
Db 2736 TNMTSLQKDLQSVRDHLAAEAKELSLILEKEDETEVQESKKACMFPEPLPKLSKSIASQTD 2795
QY 1348 GT-GLENNQEQMSDNLG---EASGCSSETSVDSCGLSSQ-----SDILTT 1394
Db 2796 GTLAKISSNOTPTPLVKNAGIQTNLQSECSE-EVTEIISQFTEKIEKMOELHAAEILDM 2854
QY 1395 QORDTMHNLK-----LQOEMAELEAVLE----- 1419
Db 2855 ESRHISETELKREHYVAVOLLKEECOTLKAVIQCLRSKEVFGFYNNCFSTLDSGSDWG 2914
QY 1420 -----QHGPSNSNYP----- 1430
Db 2915 QGIYLTSHSQGFDIASGRGESESATDFFPKIKGLLRVHNEGQVLSLITESPYSDGED 2974
QY 1431 -----SIIDSSALEDLRNPEQSTSEKAVITSSOKSE----- 1462
Db 2975 HSIQOVSEPWLEERKAVINTISSKLDLITKMLQREAEVYDSSQSHESFSDWRGELLAL 3034
QY 1463 --YPISONPEGLSADKEVSADST-----SKNKEPGVERSSPSKCPSLDDRWMY 1510
Db 3035 QOVFLERSVLLAFARTLALGTGTTDAVGLNLCLEQRIQIOGVEYQAAMECLQADR--- 3091

QY 1511 HSCSGSLQNRNYSQEELI-----KYVDEVEQOLESQPHDLTETSYLPQDLEGTPYL 1564
Db 3092 -----RSLSEIQALHAQMNGRKITLKREQSEKSPQELLEYNIQKQ-----SQML 3138
QY 1565 ESGISLSDPDPSEDAPESARVGNIPSSTSALKVPQPKVAESQSPAAAHITDTAG 1624
Db 3139 EMQVELSS-----MKDRATELQE-----QLSSEKVMVAELK-SELAOTKLELFTTLKAQ 3186
QY 1625 YNAMEE-----SVSREKPELTASTERVNKM 1650
Db 3187 HKHLKELAEARLEVKDKTDEVHLLNDTFLASEQKSKSRELQWALEKAKLGRSEERDKEEL 3246
QY 1651 SMVVGUTPEEFMLVYKPAKHHITLNLITEETHVVMKTDABFYCERLT 1701
Db 3247 -----EDLKFSLESOKORNQL-NULLQOQKLLNESQOKIESQRM 3287
RESULT 10
CENP_HUMAN
ID CENP_HUMAN STANDARD; PRT; 3210 AA.
AC P49454; Q13246; Q13171;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CENP-F kinetochore protein (Centromere protein F) (Mitosis) (AH
DE antigen).
DE CENPF.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=95348175; PubMed=7542657;
RA Liao H., Winkfein R.J., Mack G., Rattner J.B., Yen T.J.;
RT "CENP-F is a protein of the nuclear matrix that assembles onto
RT kinetochores at late G2 and is rapidly degraded after mitosis.";
RL J. Cell Biol. 130:507-518(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95379848; PubMed=7651420;
RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
RA Jones D., Yang-Feng T.L., Lee W.-H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
RT that is specifically involved in mitotic-phase progression.";
RL Mol. Cell. Biol. 15:5017-5029(1995).
RN [3]
RP SEQUENCE OF 2194-3210 FROM N.A.
RX MEDLINE=95336446; PubMed=7612011;
RA Li Q., Ke Y., Kapp J.A., Fertig N., Medsger T.A. Jr., Joshi H.C.;
RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal
RT domain sufficient for nuclear localization.";
RL Biochem. Biophys. Res. Commun. 212:220-228(1995).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95370296; PubMed=7642639;
RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
RT "The C terminus of mitosis is essential for its nuclear localization,
RT centromere/kinetochore targeting, and dimerization.";
RL J. Biol. Chem. 270:19545-19550(1995).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
CC -!- FUNCTION: PROBABLY REQUIRED FOR KINETOCHORE FUNCTION, INVOLVED IN
CC CHROMOSOME SEGREGATION DURING MITOSIS. INTERACTS WITH
CC RETINOBLASTOMA PROTEIN (RB), CENP-E AND BUBR1.
CC -!- SUBUNIT: HOMO- OR HETERODIMER.
CC -!- SUBCELLULAR LOCATION: NUCLEAR MATRIX (BUT NOT IN THE NUCLEOLUS),

REORGANIZATION TO THE KINETOCORE/CENTROMERE (CORONAL SURFACE OF THE OUTER PLATE) AND THE SPINDLE DURING MITOSIS.
-I- DEVELOPMENTAL STAGE: GRADUALLY ACCUMULATES DURING THE CELL CYCLE.
-I- PTM: HYPERPHOSPHORYLATED DURING MITOSIS.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U19769; AAA82889.1; -
EMBL; U30872; AAA82935.1; -
EMBL; U25725; AAA86889.1; -
Genew; HGNC:1857; CENPF.
MIM; 600236; -
InterPro; IPR001230; Nucleyl_site.
KW Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.
FT DOMAIN 14 197
FT DOMAIN 273 769
FT DOMAIN 823 1328
FT DOMAIN 1642 1746
FT DOMAIN 1862 2987
FT DOMAIN 2207 2568
FT REPEAT 2207 2386
FT REPEAT 2389 2568
FT DOMAIN 3015 3032
FT CONFLICT 16 16
FT CONFLICT 250 250
FT CONFLICT 272 272
FT CONFLICT 611 611
FT CONFLICT 1494 1589
FT CONFLICT 1611 1611
FT CONFLICT 1811 1811
FT CONFLICT 2242 2243
FT CONFLICT 2335 2335
FT CONFLICT 2492 2492
FT CONFLICT 2545 2551
SQ SEQUENCE 3210 AA; 367589 MW; 11D83324960E4334 CRC64;
Query Match 3.4%; Score 323.5; DB 1; Length 3210;
Best Local Similarity 19.2%; Pred No. 5.6e-06;
Matches 392; Conservative 290; Mismatches 735; Indels 621; Gaps 93;

QY 9 EYQVNVINAMQ-----KILECPICILE-----LIKEPVSTKC----- 39
DB 288 QELRNKINLELRLOGHEKEMKGOVNFQFQLOLEKAKVELIEKVKLNKCRDELVRTT 347
QY 40 ---DHIFCFCKMLLNQKKGFSQCPCKNDITKRSLOESTRFS-----OLVBEEL 86
DB 348 AQYDQASTRYTALQELKK-----LTEDLSQRQNAESARCSLEQKIKEKEFEQEL 400
QY 87 LKTIICAFQ-----LDTGLEYA-NSYNFAKKE-----NNSPEHLKD-- 120
DB 401 SROORSFQTLDOECOMKARTQELQQAKNHNVHQAELDLTQSVKQOLNLEEFQKL 460
QY 121 -----EVSIIQSMGYRNRAKRLLOSEPENPSLOETSLSVLSNLGTVRTLRT 167
DB 461 CRAEQAFQASQIKENELRSMEEKNNLLKSHSEQAREVCHLEAKNI----- 512
QY 168 KQRIQPKQTSVIEIGSDSDTQVNAKVCYSGVQDELQIQTQGTDRDISLSAKKAACE 227
DB 513 KOCLNQSQNFAEMAKNFTSQETMLRDLOEQKINQOE-----NSLTLEKLKLAVAD 562
QY 228 F-----SETDVTNTEHH-OPSNNDINTTEKH-----AARHPKPYOGSSVSNLHVEPC 274
DB 563 LEKQRCSDOLKKEHHIEQLNDKLSKTEKESKALLSALELKKKEYE-----ELK 613
QY 275 GTNTHASSLOHENSILL--LTKDRMNVEK-----ABFCNKSQOPGLARSQHNWAGSBETC 328

DB 614 BEKTLFCWKSSEKXLLQTMSEKENLQSKINHLETCLTKQO--IKSHEYNERVRLTEM- 670
QY 329 NDRRTPTSTE-----KKVDLNAADPLC-----ERKENWKQK-----LPCSENP 364
DB 671 -DRENLSVEIRNLHNVLDKSVSEVETOKLAYMELQAKAFSDQKHQKETEINCLTKTSOLT 729
QY 365 RDTEDVPWITLNSISQKYNWFSDRDELGLGSDSDHGESENAKVADVLDVLENVDE--Y 422
DB 730 QGVED-----LEHKQLQL-----SNEIMDKRCY-----QDLHAYESLRDLKSKDASLV 775
QY 423 SGSSKIDLLASDPHEALICKSERVSKSVESNIEDKIFGKTYRKKASLPNLSHVTENLI 482
DB 776 TNEHDORSLLAFDQOPAM-----HHS-----FANIIGEQSMPs----- 809
QY 483 IGAFVTEPQIITQERPLTNKLRKRRTPSGLHPEDFIKADLAVQKTPMINOGTNOTEQN 542
DB 810 -----ERS-ECRLEADQSPKNSAILQNRVDSLEFLESQOKMNSDLQKQCE-- 854
QY 543 GOVMNITNSGHENKTKGDSIONE--KNPNPTESLEKE-SAFKTKAEPITSSSISNMELEL 598
DB 855 -ELVOIKGEIEENLMKAOMHOSFVAETSQRIKQEDTSAHONVVVAETLSALENKEKEL 913
QY 599 NIHNSKAPKNRLRRKSTRIHALELVYRNLSPPNCTELQIDSCSSSEEEKKKKYNOM 658
DB 914 QLLNDKV-----ETEQAIEQLKKSNNHLEDSLEKELQ--LSETLSLEKKEMSSI 961
QY 659 PYVHRNLO-----LME-----GKEPATGAKSKNPNQETSKRHD-----SDTFPEL 700
DB 962 ISLNKREIEELTQENGTLKEINASLNQEKMNLIQKSESFANYIDREKSISELSDQYKOE 1021
QY 701 KLTNAPGSFTKCSNTSELKEFVNPSL-PREEKEELEV----- 738
DB 1022 KLI-----LLQCEETGNAYEDLSQYKAAQEKNSKLECLLNCTSLCENRKNLEBOLKEA 1077
QY 739 -----KVSNNAEDPKDMLSGERVLOQTERS-----ESSISLVPGTDYGTQESIS 784
DB 1078 FAKHQBELTFLAPAEERNQNLMELETVQOALRSEMTDNQNNKSSEAGGL--KQEIWT 1134
QY 785 LLEVSTLCKAKTEPKNKCVCQAAEFENPKGLIHGC-----SKDNRNDTEGPKYPLGHEVNHIS 840
DB 1135 LKEEO--NKMKEVNDLLOENBQMLKVMKTKHECONLESEPIRNSVKE-RESERNQCNFK 1191
QY 841 RETSTEMEESLDAQYLQNTFKVKSRQSFALFSPNPGNAEEECATFSAHSGSLKSKSPKVT 900
DB 1192 POMDLYVKEISLD-----SYNAQLVQLEAMLRNK-----ELKLOESEKE 1230
QY 901 FECEQKEENQCKNE--SNIKPVQTVNITAGFPVYVQKQKDPV----- 939
DB 1231 KECLQHELTQIRGDLTSLNLDQMOSQETS-----GLKDCEIDAEEKYISGPHELSTSQN 1284
QY 940 DNAK-----CSIKGSRFCCLSSQFRGNETGLITPNK----- 970
DB 1285 DNAHLQCSLQTTMKNKLELEICEILOAEKYELVTELDNDSRSECITATRKMAEEVKKLLN 1344
QY 971 -----HGLL-----QNPYRIPLPIKSFVK--TKCKKNL----- 998
DB 1345 EVKILNDDSGLLHGLVELDIPGGEFGEQNPQHVSLAPLDESNSYEHLTLSDKREVQMHF 1404
QY 999 --LEENF-----EHSMSPSR-----EMGNENIPSTVTSIRN----- 1029
DB 1405 AELQEKFLSLQSEHKILHDQHCQMSKMSSELQTYVDSLKAENL--VLSTNLNRFQGDLYK 1462
QY 1030 -----NIRENVEKEASSNINEVGSTNEVGSS-----INEIGSSD--ENTQAEIQRNG 1077
DB 1463 EMQGLGLEGLVPSLSSSCVDP-SSLSLSGDSFSFYRALLETQTDGMSLLSNLECAVSANO- 1520
QY 1078 PKLNAMRLGLVQPEYVKQSLPGNSKHPKTKQYEEVQTVN--TDFSPYLISD--NL 1133
DB 1521 -----CSVDEVFCSSL-----QTYVDSLKAENMLVSTNLNRFQGDLYKEMQLGL 1564
QY 1134 BQPMGSSSHASQVCSSETPDDLLDQGEIKEDTSAFENDIKES----- 1173

Db 1165 EEEVLPLSSS-CVPDSSSLSSG-----DSFYRALLEQTQDMSLNLGSLGWSANQCSV 1619
QY 1174 SAVFSKSVQKGLSRSPSPFHT-----HLAQYRRGAKKLESS----- 1212
Db 1620 DEVFCSSIQEENLTKETPSAPAKVGELESCEVYQSLKLEEKMESOGIMKNKBIQE 1679
QY 1213 -EENLSSEDELPQHLFGKVNIPQSQT-----RHSTVATECLSKNTEENLLSLKNS 1266
Db 1680 LEQLLSERQELDLRKQYLSENQWQKLTSVTLEMESKLAAB--KKQTEQLSLELEVA 1737
QY 1267 L-----NDCSNQVTLAKAQEHLLSETKCSASFQSCSELEDLTANTQDFLIGSS 1321
Db 1738 RLQQLGLDLSRSLGIDTDAIQRNESCDIS-----KEHTSETTERTP----- 1782
QY 1322 QOMRHQSESQVGLSKDELVSDD--DEER--GTGLEENNOE---EQSMDSNL--GE--- 1367
Db 1783 KHDVHQ-----ICDKDAQODLMDIEKITETGAVKPTGECSEQSPDINYEPPGEDKT 1835
QY 1368 -AASGCSETSVS-----ECSGLSSQSDILTTOOR--DTMQHNL-----IK 1406
Db 1836 QGSSECISELFSFGPNALVPMDFLGNQEDIHNLQLRKETSNEMLLHVIEDRDKVES 1895
QY 1407 LQQEMAELEAVLEQHGQSPNSYPSIISDSSALEDLRNPQSTSEKAVLTSQKSE-YPI 1465
Db 1896 LLNEMKELDSKLHQEVQLMTKIEACIELEKIVGELKKENDSUSEKLEIFYFSCDHQELLQR 1955
QY 1466 SONPEGLSADKFEVSADSSSTKKNKEPQGVERSSPKCPSLDDRWMHSCSGSLQNRNYPQ 1525
Db 1956 VETSEGLNSD-LEMHAQKSSREDIGDNVAK-----VNSDW-----KERFLDVE 1997
QY 1526 ELIKV-----VDVEEQLESGPHDLTFETSYLPRDLEGTPYLESGISLFSDDPESDPE 1581
Db 1998 NELSIRSEKASIEHEALYLEADLEVQVTEKL---CLEKDNENKQKIVIVCLEEELSVMVS 2054
QY 1582 DRAPESARVGNIPSTSAIKVPQLKVAESAQSPAAATTTDTAGYNAMESVSREKPEL 1639
Db 2055 ERNOLRGELDTMSKTTALDQLSKMEKTEKTOE-LESHQSECLHCIOVAEAEVREKTEL 2111

RESULT 11
YM67_YEAST
ID YM67_YEAST STANDARD; PRT; 1658 AA.
AC Q03661; Q04988;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 187.1 kDa protein in GUA1-ERC8 intergenic region.
GN YMR219W OR YMR261.13 OR YW9959.01.
OC Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-711 FROM N.A.
RC STRAIN=S288c / AB972;
RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 608-1648 FROM N.A.
RC STRAIN=S288c / AB972;
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 249809; CAA89934.1; -.

DR EMBL; 249939; CAA90190.1; -.
DR SGD; S0004832; YMR219W.
KW Hypothetical protein.
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match 3.3%; Score 322; DB 1; Length 1658;
Best Local Similarity 19.7%; Pred. No. 2.8e-06;
Matches 344; Conservative 255; Mismatches 704; Indels 444; Gaps 76;

QY 132 NRAKRLLOSPEPNPSLOETSLSVQLSNLGTVRLTKQRIQKQTSYVIELGSDSSSDTV 191
Db 97 NRGSILTTLEKEDALFERSLEEE-----RQRFQ-----LHDSLM 131
QY 192 NKATYCSVGQQLLI--TPQGT-----RDEISLSAKKAACEFSFSDTNTNTEHIQPSN 243
Db 132 NKYTGNKSQHLIDLRLKSOYQDTFTSQNNDEIPLDS-----FISPLPDAEDRESSN 184
QY 244 NDLNTEKRAERHPEKIQGSSVSNLHVPCGNTNTHASSLQHENS-----SLLLTNDR- 296
Db 185 IDSDDKDELEGKQ-----SLIKDFLE---NDEVELSEEEKNSGQSSPSIMILSDEE 234
QY 297 -----MNVEKAFCNKSQOPGLARSOHN-----RWAGSKETCNDRRTPSTTEK 339
Db 235 YABEGALQDVSNDYEAEQGVKRNIGQEQANVENATQISSDSSSEGQNYSEGVEMELE 294
QY 340 VDLNADPLCRKE-----MNKQLPCSENPRDTEDPWI-TLNSSSIQKYNWF 386
Db 295 DDIDVESDAEKDESOGAEGTEHSVDFSKYMQPRTDNTK----IPVTEKYESDEHKVHQRY 350
QY 387 SR-----SDELLGSDSDHGESESNAKVADVLVDLNEVDYSGSSEKI-DLLASDPHE 438
Db 351 SEDGAFDGVSNISVDDDESEDAESYASANAENVVHHNEHEDLDDKELIEDIESD--- 407
QY 439 ALICKSERVHSKSVESNIEDKIFCKYIRKASLPNLSHVTEENLLIGAFVTEPOIIQERPL 498
Db 408 -----SESQAQSEQSGEDDFEYKMKNEKSTSEETNTSES-----RDQGFADKDAYT 455
QY 499 TNKLKRRRRTSGTLHPEDFIKKADLAVQKTPMINQCTNOTEQNGOVNMNTNSGHENKT 558
Db 456 KNKVEQGENDE--EPE-----KDDIRSSLDKNPHGNKNSKEYSENVL-----ENETD 501
QY 559 GDSIQNEKNPNPIESLEKESAFKTKAEPISISSISNMELELNIH----- 601
Db 502 PAIVERENQINDVEGYD-----VTGKSVESDLHEHSPDNLVLAARAMLQ 546
QY 602 -----NSKAPKKNLRLRKSSTRIHALELVVSRNLPPNCTELQIDSCSSSEIKKKYN 656
Db 547 FQOSRNSNCPQKE-----EQVSESYLGHSGNSLNGSLRSLDESEE----- 585
QY 657 QMPVR-----HSRNLQIMECKEPATGAKKSNKNEQTSKRHDSDTPELKLTNAPGSFTK 711
Db 586 QIPLKQFTGNNNNLKTDRGLDSSSVIEVEKYSE---KKLDGSTKEKEL----VPLSTD 638
QY 712 CSNTSEL--KEFVNPSLPREEK-EEKLETVKVYNNNAEDPKDMLMSGERVLOTERSVESS 768
Db 639 TINSSLGNEDSIYSLDDADAISENLTDPVLMETTPKRYEIVIESV-YSSTSYEDNT 697
QY 769 ISLVP-----GTDYGTQESISLLEVSTLGLKAKTEPNKCVSQCAAFENPK 812
Db 698 VAMPPQVEYTSPPFMDNPFNSLNDDEYKKHLLKSTLAALAPAFTKKD-----AEEVE-A 750
QY 813 GLTHGSGKDNRRNTEGFKYPLGHEVNHRSRETSIEMESSELDQY-LQNTFKVKSKROS--F 869
Db 751 GVTKSLTSTSGHTNIF-----HTSKETKQVS-DLDESTENVTFENGTGDKNKSKNF 804
QY 870 ALFSNPGNAEEECATPSAHSGSLKQSPKVTFECEQKEENQGNESNI-----KPVQTVNI 925
Db 805 PGVANSTDKSTEDNTDEKYFSAINVTNVTGSSCDIEITASNVEENLRYCEKDMNEAEM 864
QY 926 TAGFPVVGQDKPVDNAKSIKSGSRFLCSS--QFR-GNETGLITPNKHGLLONPFRIPP 982
Db 865 SSGDECYKQND---DGSKTQIS----FSTDSPDNFQESNDNTEFSSTKYKV-----RNSD 912

DB 4259 PEAKTK 4264

RESULT 13

ATRX_HUMAN

ID ATRX_HUMAN STANDARD; PRT; 2492 AA.

AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;

DT 01-NOV-1995 (Rel. 32, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked nuclear protein) (XNP) (Znf-HX).

GN ATRX OR RAD54L OR XH2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANT S-1860, AND VARIANTS ATR-X.

RX MEDLINE=97123494; PubMed=8968741;

RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J., Gibbons R.J.;

RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome.";

RL Hum. Mol. Genet. 5:1899-1907(1996).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).

RX MEDLINE=97386582; PubMed=9244431;

RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P., Colleaux L., Schwartz C., Fontes M.;

RT "Determination of the genomic structure of the XNP/ATRX gene encoding a potential zinc finger helicase.";

RL Genomics 43:149-155(1997).

RN [3]

RP SEQUENCE OF 860-2492 FROM N.A.

RX MEDLINE=95179111; PubMed=7874112;

RA Stayton C.L., Dabovic B., Gulisano M., Geetz J., Broccoli V., Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E., Bianchi M.E., Consalez G.G.;

RT "Cloning and characterization of a new human Xql3 gene, encoding a putative helicase.";

RL Hum. Mol. Genet. 3:1957-1964(1994).

RN [4]

RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.

RX MEDLINE=94214473; PubMed=8162050;

RA Geetz J., Pollard H., Consalez G., Villard L., Stayton C.L., Millasseau P., Khrestchatsky M., Fontes M.;

RT "Cloning and expression of the murine homologue of a putative human X-linked nuclear protein gene closely linked to PGK1 in Xql3.3.";

RL Hum. Mol. Genet. 3:39-44(1994).

RN [5]

RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.

RX MEDLINE=95211835; PubMed=7697714;

RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;

RT "Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome).";

RL Cell 80:837-845(1995).

RN [6]

RP SEQUENCE OF 1375-2492 FROM N.A.

RA Pearce A., Chapman J.;

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

RN [7]

RP E2H2 BINDING.

RX MEDLINE=98167853; PubMed=9499421;

RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M., Colleaux L.;

RT "Specific interaction between the XNP/ATRX gene product and the SET domain of the human E2H2 protein.";

RL Hum. Mol. Genet. 7:679-684(1998).

RN [8]

RP SURCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC HETEROCHROMATIN.

RP

RX MEDLINE=20040663; PubMed=10570185;

RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M., Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J., Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;

RT "Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes.";

RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).

RN [9]

RP DISEASE.

RX MEDLINE=20213147; PubMed=10751095;

RA Villard L., Fontes M., Ades L.C., Geetz J.;

RT "Identification of a mutation in the XNP/ATRX gene in a family reported as Smith-Fineman-Myers syndrome.";

RL Am. J. Med. Genet. 91:83-85(2000).

RN [10]

RP VARIANT ATR-X SER-1713.

RX MEDLINE=97196774; PubMed=9043863;

RA Villard L., Lacombe D., Fontes M.;

RT "A point mutation in the XNP gene, associated with an ATR-X phenotype without alpha-thalassemia.";

RL Eur. J. Hum. Genet. 4:316-320(1996).

RN [11]

RP VARIANT JM GLN-2131.

RX MEDLINE=96224392; PubMed=8630485;

RA Villard L., Geetz J., Mattei J.-F., Fontes M., Saugier-Verber P., Munnich A., Lyonnet S.;

RT "XNP mutation in a large family with Juberg-Marsidi syndrome.";

RL Nat. Genet. 12:359-360(1996).

RN [12]

RP VARIANTS ATR-X.

RX MEDLINE=97467722; PubMed=9326931;

RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asenbauer B., Bergoffen J., Berry S.A., Dahl N., Fryer A., Reppier K., Kurosawa K., Levin M.D., Masuno M., Neri G., Pierpont M.E., Slaney S.F., Higgs D.R.;

RT "Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain.";

RL Nat. Genet. 17:146-148(1997).

RN [13]

RP VARIANT ATR-X LEU-246.

RX MEDLINE=20123062; PubMed=10660327;

RA Fichera M., Romano C., Castiglia L., Falla P., Ruberto C., Amata S., Greco D., Cardoso C., Fontes M., Ragusa A.;

RT "New mutations in XNP/ATRX gene: a further contribution to genotype/phenotype relationship in ATR/X syndrome.";

RL Hum. Mutat. 12:214-214(1998).

RN [14]

RP VARIANT SHS LYS-1742.

RX MEDLINE=99347960; PubMed=10417298;

RA Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C., Prieto F., Fontes M., Martinez F.;

RT "Mutation of the XNP/ATRX gene in a family with severe mental retardation, spastic paraplegia and skewed pattern of X inactivation: demonstration that the mutation is involved in the inactivation bias.";

RL Am. J. Hum. Genet. 65:558-562(1999).

RN [15]

RP VARIANT CWS THR-2050.

RX MEDLINE=99326061; PubMed=10398237;

RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M., Curtis M.;

RT "Carpenter-Waziri syndrome results from a mutation in XNP.";

RL Am. J. Med. Genet. 85:249-251(1999).

RN [16]

RP VARIANTS ATR-X E-175; 178-V--K-198 DEL; S-190; P-219; L-246 AND C-249.

RX MEDLINE=99219535; PubMed=10204841;

RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Belougue J., Lossi A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M., Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;

RT "Evaluation of a mutation screening strategy for sporadic cases of ATR-X syndrome.";

RL J. Med. Genet. 36:183-186(1999).

Db 720 TVSVISDTSQKLAIILOEHLHKEQEAWEATSEDSVVIIVDVKEPSRVDVSCPELEGVE 779
QY 690 KRHSDTFPELKLITNAPGFTKCSNTSELKEFVNPSLPREEKEKLETKVKSNNADPK - 748
Db 780 KCSDSQSWEDI-----APEITEPCAENRLDTKEEKSVEYEGDLKSGTAETEPVEODSSQPSL 835
QY 749 -----DLMLSGERVLTGTERSSESSISLVPDGYDGTQESISL-LEVSTLTGAKTEPNKCV 802
Db 836 PLVRADDPRLRLQLOEQOQTOEKTNSLTEDSKMANAKOLSSDAEAQKLGKPSAHASQSF 895
QY 803 SOCAA-----FENPKGLIHGCKDNNDTEGFKYP-----LGH-----EVNHSREIS 844
Db 896 CESSSTPHFTLPK-----EGDIIPLTGATPPLIGHLKLKPRKHSTPTIG 941
QY 845 IEMEESELDAQYLQNTFKYKRSQSFALFSN-----PGNAEEECATFSAHS 889
Db 942 I-----SNYPESTIATSDVMSESMVETHDPIILGSGKGDSCAAAPDVKDKLCLRMKLV 993
QY 890 GSLKKSPPKVTTECEKEBNOGKN-----ESNIKPVQVNTITAGPPVVGQKD--KPDVN 941
Db 994 PETEASEESLOFNLEKRPATGERKNGSTAVAESVSPQKTMVSLSCICEARQENEARSEDP 1053
QY 942 AKCSIKGSRFCLSQFRNETCLITPNKHGILLONPYRIPPLFPKISFVYTKCKK----- 996
Db 1054 PTPPIRGNLLHFPSSQ--GEEKEKLEGDHTIRQSQOPMKPISPVKDPVSPASQKMVIQ 1111
QY 997 -----NLEENFEHNSPSPER-----EMGEN-----IPSTVSTIS 1027
Db 1112 PSSPQGEAMVTDVLEQKEGRSTNKENPSKALLERSQNNIGIOTMECSLRVPETVSAAT 1171
QY 1028 RNNIRENVKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNRPKLNAMLRLG 1087
Db 1172 Q-----TIKNVCGQSTVD-----QNFGRQDQATVQTERSGEKP----- 1206
QY 1088 VLQPEVYKSLGCSNCKHPEIKQEEVEVQVNTDFSYLISDNLEQPMGSSHASQVCS 1147
Db 1207 -----VSAQGDDE--SLHSQGGEEF-----DMPQPPHG--HVLURHM 1240
QY 1148 ETDPDLDLDDGEIKEDTSFAENDIKESSAVFSKSVQRGELSRSPPTHTHLAQGYRRGAK 1207
Db 1241 RT-----IREVRTLVTRI-----TDVYVYDGTVEVERK 1268
QY 1208 KLESSENLSSDEEULPCFQHLLFGKVNIPQSOSTRHSTVATECLSKNTEENLLSKNLSL 1267
Db 1269 VTEETEPEPT-----VEQCEQETEVSQOTGGSS 1296
QY 1268 NDCSN-QVILAKASQBHLSEETKCSALFSSOCS-----ELEDLTANTNTQDPFLIGSS 1321
Db 1297 GDLGDISSFSSKASSLHRTSSGTSLSA-MHSSGSGKGAGPLRGKTSGTPEPADFALPSSR 1355
QY 1322 KQMRHQSESGVGLSDKELVSDDEERGTLGLENNQBEQSMDSNLGEAAGCESETSVSSED 1381
Db 1356 GPGKLSPRKGVSYQTGTPVCEEDGDAGLIRQ-----GKAPVTTR 1396
QY 1382 CSGLSQSQDILTTQORDT-----MOHNL-----IKLOQEMAELEAVL 1418
Db 1397 GRGRGRPSRTGTRETATVPGLGTIEDISPNLSPDKSFSRVVPRVPDSTRRTDVGAGA 1456
QY 1419 EOHGSPSNYSPIISDSSALEDLRNPQOSTSEKAVLTSQKSEY----- 1463
Db 1457 LRSDSPETPFOAAGPSDGL-DASPGMSFVGLRVKAKWSNGSNYPYSGKITRVDGAGKY 1515
QY 1464 -----PISQNPGLSADKFEYSADSSTSKNKEPG-----VBR 1495
Db 1516 KLLFDDBGYECDVLGDKDILLCDPIPLDTEVTALSEDEY-FSAGVVKGRKESGELYYSIEK 1574
QY 1496 SPSKCPSLDDRWYH-SCSGSLQNRN-----YPSQBELIKVVDVEEQLEE--- 1541
Db 1575 EGQRK-----WYKRMVILISLEQGNRLREQYGLCPYEAVTPLTKAADISLDNLVEGRK 1627
QY 1542 -----SGPHDLTETSYLRQDLEGTPLYESGI-----SLFDDPDSPESDRAPES 1587
Db 1628 KRNSVNSPATTSASSSTTPTTKITESPASMGVLGSKRKLITSEERSPAK-RGRKS 1686

QY 1588 ARV-----GNIPSSSTSALKVPQLKVAESAQSPAHAHTTDTAGY----- 1625
Db 1687 ATKVPCAVGAGEFVSPCESGDNTEBPSAL-----EQRGPLPLNKTLFLGYAFLITM 1738
QY 1626 -NAMESVSREK-PE-LTASTERVNRKMSVMVSGLTPEEFMLVYKARKHHTLTNLITE 1682
Db 1739 ATTSDKLASRSLPDGPTGSSEE-----EEEFLEIPPENKQY----- 1775
QY 1683 ETHVVMKTDADFVCP-----RTLKYFLGTAGGKVVVSYFVFWTQ 1721
Db 1776 --TESQLRAGAGYILEDNFNEAOCNTAYOCLLIADQHCRTRKFLCLASGIPCWSHVWV-- 1831
QY 1722 SIKERKMLNEHDFEVRGDVNVGRNHQGP-----KRARESDQRKIFRGLIEICCCYGPFTN 1774
Db 1832 -----HDSCHANQLQNYRNYLLPAGYSLEEQRIILDWQPRE-----NPFQ 1871
QY 1775 MP-----TDQLEWVQLC-----GASVYKEIJSFT-----LGTGVHPVIVVOPDAWTE 1819
Db 1872 LKVLVLSVDOQQNFLELWSEILMTGTGAASVKQHHSSAHNKDIALGVDPVVVTDPSC---PA 1928
QY 1820 GFHAIGOMCEAPVVTREWVLDSVALYQC 1847
Db 1929 SVLKCAEALQLPVVSOEWVI-----QC 1950

RESULT 15
ALM1_SCHPO
ID ALM1_SCHPO STANDARD; PRT; 1727 AA.
AC Q9UTK5; O13313; Q9UTU8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Abnormal long morphology protein 1 (Sp8).
GN ALM1 OR SPAC1486.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Guilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Welljones I., Volktaert G., Aert R., Robben J., Grymonprez B.,
RA Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomycetes pombe.";
RL Nature 415:871-880(2002).
RN [2]
RP SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.
RC STRAIN=972;

RX MEDLINE=20123449; PubMed=10660053;
 RA Jimenez M., Petit T., Gancedo C., Goday C.;
 RT "The *alm1+* gene from *Schizosaccharomyces pombe* encodes a coiled-coil
 RT protein that associates with the medial region during mitosis.";
 RN Mol. Gen. Genet. 262:921-930(2000).
 RP [3]
 RC SEQUENCE OF 644-834 FROM N.A.
 RX STRAIN=968 h90;
 RA Ding D.Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
 RA Hiraoka Y.;
 RT "Large-scale screening of intracellular protein localization in living
 RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";
 RL Genes Cells 5:169-190(2000).
 CC CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN
 CC CYTOKINESIS.
 CC CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN
 CC CYTOKINESIS.
 CC SUBCELLULAR LOCATION: Cytoplasmic.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL133357; CAB62414.1;
 DR EMBL: AF010473; AAB65416.1; ALT_INIT.
 DR EMBL: AB028012; BAA87316.1; -;
 FT Coiled coil.
 FT DOMAIN 57 361 COILED COIL (POTENTIAL).
 FT DOMAIN 443 463 COILED COIL (POTENTIAL).
 FT DOMAIN 542 740 COILED COIL (POTENTIAL).
 FT DOMAIN 804 1106 COILED COIL (POTENTIAL).
 FT DOMAIN 1223 1427 COILED COIL (POTENTIAL).
 FT DOMAIN 1497 1555 COILED COIL (POTENTIAL).
 FT DOMAIN 1601 1664 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1727 AA; 197858 MW; F820BF8D9C132644 CRC64;
 Query Match 3.38; Score 316; DB 1; Length 1727;
 Best Local Similarity 20.18; Pred. No. 5.2e-06;
 Matches 369; Conservative 267; Mismatches 652; Indels 548; Gaps 84;
 QY 7 RVEEYQVNIAMQKILEPICLELIKPEVSTKCDHFICMCLKLQ----- 54
 DB 85 QAEERNLISVLRN-----ENSLKTNLENQNKRFDAITENQSLRRANSELOEQ 134
 QY 55 -KGPSQCPCKNDITKRSLO-ESTRFSQVVEELLKTIICAFQDGTGLEAYANSYNFAKKEN 112
 DB 135 SKIASEQLSIKADQI--EALQNENSHLGEQVQSAHQALSIDIE-----ERKKQHMFA---- 183
 QY 113 NSPEHLKDEVSIIQSMGYRNRKRLLOSEP-----ENSLQETSLSVOLSNLGTVRT- 164
 DB 184 SSSSRVKEEI-LVQEKSAVSALSDASLDHSKVGCEKLEVSRSQVQDLKLAGLAQONTE 242
 QY 165 LRTKRIQOPKTSVYIELGSDS-----SEDVTNKATYCSVGDOELQITPGTDRDEISLDS 220
 DB 243 LNEKIQLEQKRSYSDGNISLKTLEDTPTSIKELEEVEVETQKRLTALWESKSELOSEV 302
 QY 221 AKKAACESETDVTNTEHHQPSNN--DLNTEKRAARHPKPKYGGSSVSNLHVPCGTNT 278
 DB 303 AALQELTSSQSLNNVTEELNNKQKLLISENSLRELQ-EKYD-SVVSSELOVVKENKNT 360
 QY 279 HASS-----LQENSSLLLTDKDRMNVKEAFCNK 307
 DB 361 SVSAGVGLFSLAQLSAVQNPESFTKVYSDNNKLOQKVSLSKLQDLRLTNKFSFCEQ 420
 QY 308 SKQ--PGLARSOH-----NRWAGSKETCNDRRPTSTPEKKVDLNADPLCERKE---- 352
 DB 421 VKQRPVVKQORSEIVRNINYNFLESLETSNNLT-----KVQAEELSTKMRQECAYL 475
 QY 353 -WNKQKLPCESENPRDTEVPWITLNSSIQKYNWFSSRS-----DELLGSDDDSHDGE 402

Db 476 QLTASRTQCSDLSEV-----ICLMAELDLHNETKSRNPATVQVALDE-YAQNSTASE 529
 QY 403 SESNAKVADVLVDVNEVEYSGSEKIDLLASDPHEALICKSERVHKSVSNSIEDKIFG 462
 Db 530 TLNVKELANFSSIKEAVSKTLELREKV-----RALECDVE-IQKQTVQYQISNAVKE 580
 QY 463 KTYRKKASLPNLHVTEENLIIGAFVTEPQIIQERPLTNKLRKRRP-----TSLGHPED 516
 Db 581 NSNLTSEQIKNL-----ESELNSSKIKNESLNLNRLKEMLATSRSSILSHNSSAGNIDD 636
 QY 517 FIKKAD-----LAVQKTPEMINQOTQNGQVMNITNSGHENKTKG 559
 Db 637 KMKSIDESTRELEKNYEVYRNMETAIQESLSKRNQDL-LSEMEALRKELENSKYQOQLST 695
 QY 560 DSTONENKPNPIESLEKE-----SAFKTKAEPISSSISNMELE 597
 Db 696 DRLTNAN--NDVEAFKKKEAKELRSINQNLQDIIISQDORASKFABELHLVNSLAERLKE 753
 QY 598 LNIHNSKAPKKNRURRKSSTRHIALE-LVVSRLNSPPNCTELO-IDSCSSSEIKKKK 654
 Db 754 LNA--SKGKD--LRKTRQERLISENCKLLAERELMSVLDLQTFLLNQOQLSDAARKVK 809
 QY 655 YNQMPVHRHRLQMEGKEPATGAKSN-----KPNK 686
 Db 810 FES--EKESLSLSQKLKE--SNEKMSNDLHSLQSKLEKSGIEYSSRIKTLMLKQKSLSE 865
 QY 687 QTSKRHDSDTFPELKLTNAPG-----SFTKCSNTS----- 716
 Db 866 DNRKLLDNQOMMEIKTLOELNGVIELEKORFTLEAKFTQOKNTSYSEREALLESLSLQ 925
 QY 717 ----ELKEFYVNPSPREEK-----EEKLETYKVSNNAEKPKDMLSGERVQIQTESRV 764
 Db 926 SKHTSLESQYNSLRNIEQLOAAASKLAEEWVERVKTEVD-----EYRLOTSESL 974
 QY 765 ESSSISLVPGTDXGTQBSISLLEYSTLGKARTPEKVCVQCAAFENPKGLIHGSKDNRN 824
 Db 975 EKNHLKIT-----SLEQRIVILQ-----DEIASSSLRCEN-----ITKDET 1011
 QY 825 DT-----EGPKYPLGHEVNHVSRETSIEMSESELDQAYIQNTFKVSK-----RQSFALFNP 875
 Db 1012 RVALLLEENKH-LNNEUSSHR--NAEKQHLKENDYKQQLLLVTDLRLKTRDY----- 1062
 QY 876 GNAEECATFSAHSGSLKQSPKVTFECEQKEE-----NOGKNESNIKPVQTVNITAG 928
 Db 1063 ---EKELLRADARSTLOKLREDTYKALEQVEDLNKELALKAGINESQ-----P 1108
 QY 929 FPVVGQKDPVDNAKCSIKGGSRFCL-----SSQPRGNETGLITPN-----KHGL--LQNP 977
 Db 1109 FP-ISEKEDPLRQEVYVVKQNAAMLTLQLOSSLNLFABEITSPSPDLDSVMKGLGLDLQNH 1167
 QY 978 Y-RIPPLFPPIKSFVTKCKKNLLLEENFEHSHMSPEREMGNENISTVSTISR--NNIREN 1034
 Db 1168 VKRISKEIMEIIS-----CQQLL---FLE-----NKKLRTVSSNRVADLQRG 1209
 QY 1035 VFKEASSNNINEVGSST-----NEVGSSINIEGSSDENIQAEILGRNRPKLN---A 1082
 Db 1210 ITEKDVSTSESVGERSNYLNMVALLNESKNLSRLNERNEEVITEL-REKIELKTDLA 1268
 QY 1083 MLRLGLVQPEVYKQS-----LPGSNCKHPETIKQBYEYVQTVNTDFSPYLLSDNLEQ 1135
 Db 1269 NFRLNKEQLESQLTEKAQVKKLENSN---EYKRNHQEILLNSLSSTST----- 1315
 QY 1136 PMGSSHASQVCE--TPDDLDD-----GEIK-----EDTSAENDIKESSAVFSVKQGE 1185
 Db 1316 ---SSDASRLKLNLSKENLIEELNQLKSELETYKSKSEDLENERAQNSKIEOLE 1372
 QY 1186 LSRSPSPFTTHLAQGYR-----RGAKKLESSENLSEDEEL-----PCFQ-HLLF 1231
 Db 1373 LK-----NTKLAARWTKYEQVNVKSLKHNQIROLQSKTSELEAKVAECHQLNQL 1425
 QY 1232 GKNVNTIPSQSTRH--STVATECISKNTEENLLSLKNSLNDSCSNQVILAKASOEHLSEET 1289

Db 1426 NKPSATPTATQSEPSTVSLSEFN-STKELSSTQKLSKSEIMD----- 1467
QY 1290 KCSASLFSSQCELEDTANTQDPELIGSSK-----QMRHQSESQGVGLSDKELV 1341
Db 1468 -----ILNTTKEELEKVRQSNKSE-----GTSKDTEIPNEEEMERKKVMQOEVLRLRSRI 1518
QY 1342 SDDEERTGLEENNQEESQSDNLGEAASCESETSVS----- 1379
Db 1519 AKELQKNELLRKONQVLOQOVKALQETVVSSEAEASVHADTKDLENLKKTEEMLSVTF 1578
QY 1380 -----EDCSGLSSOSDILTT-----QORDTMOHN----- 1403
Db 1579 QVIFNESISDFSTADFTTFVQKEWEKREIILQKDVEEQVAQSHQKQLDNIRKELEMRN 1638
QY 1404 ---LKLQOEAMAELEAVLEQHGQSPNSYPSIISDSSALEDLRNPEQSTSEKAVLTQ-- 1458
Db 1639 KLKLSMLEKNLARVRAELEQSKKKDS---FALLS---LEASKNTDSNKSNSSEVPRAQVK 1691
QY 1459 -----KSSEYPISONPEGLSADK-FEVSADSSTSK 1487
Db 1692 EKXLIAKTHSVDTNSPPKRSDDAGMDVSNDRKAK 1727

Search completed: January 22, 2003, 16:50:38
Job time : 80.5 secs

GenCore version 5.1.3

Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:41:53 ; Search time 68 Seconds

(without alignments)
5645.085 Million cell updates/sec

Title: us-09-734-672-4

Perfect score: 9649

Sequence: 1 MDLSALRVEEONVINAMQK.....LYQCQLDLYLIPQIPHSY 1863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_21.*

- 1: sp-archaea.*
- 2: sp-bacteria.*
- 3: sp-fungi.*
- 4: sp-human.*
- 5: sp-invertebrate.*
- 6: sp-mammal.*
- 7: sp-mhc.*
- 8: sp-organelle.*
- 9: sp-phage.*
- 10: sp-plant.*
- 11: sp-rodent.*
- 12: sp-virus.*
- 13: sp-vertebrate.*
- 14: sp-unclassified.*
- 15: sp-rvirus.*
- 16: sp-bacteriap.*
- 17: sp-archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9471	98.2	1863	6 Q9GKK8	Q9gkk8 pan troglod
2	9233.5	95.7	1792	4 O15129	O15129 homo sapien
3	5795	60.1	1141	6 O46485	O46485 gorilla gor
4	5776	59.9	1141	6 O46484	O46484 pan troglod
5	5671	58.8	1141	6 O46486	O46486 pongo pygma
6	5443.5	56.4	1140	6 O46487	O46487 macaca mula
7	5014.5	52.0	1146	6 O46488	O46488 alouatta se
8	4895	50.6	1812	11 Q921D2	Q921d2 mus musculus
9	4873.5	50.5	1817	11 Q54952	Q54952 rattus norv
10	4386.5	45.5	1146	6 O46490	O46490 cynocephalu
11	4040	41.9	1141	6 O46489	O46489 galago cras
12	3693	38.3	1063	11 Q924E0	Q924e0 pedetes cap
13	3557.5	36.9	947	6 Q8WMT3	Q8wmt3 tapirus pin
14	3555.5	36.8	947	6 Q9BDV3	Q9bdv3 diceros bic
15	3482	36.1	942	6 Q9BDV4	Q9bdv4 equus cabal
16	3479	36.1	973	11 Q924E1	Q924e1 graphiurus

17	3452	35.8	944	6 Q9BDV2	Q9bdv2 lama glama
18	3435	35.6	959	11 Q924E3	Q924e3 sciurus nig
19	3419.5	35.4	963	11 Q99MX4	Q99mx4 glaucomys v
20	3406.5	35.3	931	6 Q9BDW2	Q9bdw2 chaetophrac
21	3382	35.1	948	6 Q9BDU6	Q9bdue6 felis silve
22	3361	34.8	942	6 Q9BDU8	Q9bdue8 physeter ca
23	3356	34.8	940	6 Q9BDU7	Q9bdue7 megaptera n
24	3345	34.7	759	4 Q92897	Q92897 homo sapien
25	3337.5	34.6	931	6 Q9BDU4	Q9bdue4 trichechus
26	3336.5	34.6	949	6 Q9TT62	Q9tt62 hipposidero
27	3318	34.4	934	6 Q8SQG2	Q8sqg2 natalus str
28	3298	34.2	946	6 Q9BDV0	Q9bdv0 sus scrofa
29	3294	34.1	938	6 Q9BDW0	Q9bdw0 bradypus tr
30	3293	34.1	924	6 Q9BDU5	Q9bdue5 dugong dugo
31	3283.5	34.0	945	6 Q9BDW1	Q9bdw1 tamandua te
32	3276	34.0	936	6 Q9TT68	Q9tt68 myotis daub
33	3263	33.8	942	6 Q9BDV5	Q9bdv5 manis sp o
34	3262.5	33.8	931	6 Q8WMT8	Q8wmt8 socialia flu
35	3245	33.6	930	6 Q9BDU3	Q9bdue3 loxodonta a
36	3236	33.5	906	6 Q8SQF7	Q8sqf7 rhinopoma h
37	3234	33.5	928	6 Q9BDU2	Q9bdue2 elephas max
38	3226	33.4	928	6 Q9BDU9	Q9bdue9 hippopotamu
39	3219	33.4	944	6 Q9BDV7	Q9bdv7 scalopus aq
40	3217	33.3	930	6 Q9TT67	Q9tt67 tadarida br
41	3197	33.1	928	6 Q9BDW3	Q9bdw3 dasypus nov
42	3171.5	32.9	917	6 Q8WMT4	Q8wmt4 roussetus a
43	3171	32.9	935	6 Q9TT63	Q9tt63 pteropus ra
44	3167	32.8	930	6 Q8SQG1	Q8sqg1 rhinolophus
45	3157.5	32.7	929	6 Q9BDT4	Q9bdt4 orycteropus

ALIGNMENTS

RESULT 1

Q9GKK8	ID	Q9GKK8	PRELIMINARY;	PRT; 1863 AA.
AC	Q9GKK8;	DT	01-MAR-2001 (Tremblrel. 16, Created)	
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)			
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)			
DE	BRCA1 (Fragment).			
OS	Pan troglodytes (Chimpanzee).			
GN	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.			
OX	NCBI_TaxID=9598;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BLOOD;			
RA	Takeda R., Hink R.L., Jogodka C., Walter N.A.R., Messier W.;			
RT	"Positive selection on the human BRCA1 gene may have resulted from			
RT	pressure for prolonged care for infants.";			
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.			
CC	!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.			
DR	EMBL; AF207822; AAG43492.1; -			
DR	InterPro; IPR001357; BRCT.			
DR	InterPro; IPR002378; Brst_cancerI.			
DR	InterPro; IPR001990; Granin.			
DR	InterPro; IPR001841; Znf_ring.			
DR	Pfam; PF00533; BRCT; 2.			
DR	Pfam; PF00097; zf-C3HC4; 1.			
DR	PRINTS; PR00493; BRSTCANCER1.			
DR	SMART; SM00292; BRCT; 2.			
DR	SMART; SM00184; RING; 1.			
DR	PROSITE; PS50172; BRCT; 2.			
DR	PROSITE; PS00422; GRANINS_1; UNKNOWN_1.			
DR	PROSITE; PS00518; ZF_RING_1; 1.			
KW	Zinc-finger. 309			K -> E.
FT	VARIANT 309			S -> G.
FT	VARIANT 590			E -> K.
FT	VARIANT 731			E -> G.
FT	VARIANT 1100			E -> G.

OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RP	[1]	
RC	SEQUENCE FROM N.A.	
RC	TISSUE=TESTIS;	
RX	MEDLINE=96172829; PubMed=8589721;	
RA	Holt J.T., Thompson M.E., Szabo C., Robinson-Benion C., Arteaga C.L.,	
RA	King M.C., Jensen R.A.;	
RT	"Growth retardation and tumour inhibition by BRCA1.";	
RT	Nat. Genet. 12:298-302(1996).	
RL	[2]	
RL	SEQUENCE FROM N.A.	
RP	TISSUE=TESTIS;	
RC	Holt J.T., Robinson-Benion C.;	
RA	Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.	
RL	EMBL; AF005068; AAB61673.1; .	
DR	InterPro: IPR001357; BRCT.	
DR	InterPro: IPR002378; Brst_cancerI.	
DR	InterPro: IPR001990; Granin.	
DR	pfam; PF00533; BRCT; 2.	
DR	PRINTS; PR00493; BRSTCANCER1.	
DR	SMART; SM00292; BRCT; 2.	
DR	PROSITE; PS01172; BRCT; 2.	
DR	PROSITE; PS00422; GRANINS_1; UNKNOWN_1.	
DR	SEQUENCE 1792 AA; 199692 MW; OCB9B06C0791A44 CRC64;	
SQ		
Query Match 95.7%; Score 9233.5; DB 4; Length 1792;		
Best Local Similarity 99.8%; Pred. No. 0;		
Matches 1788; Conservative 0; Mismatches 3; Indels 1; Gaps 1;		
QY	72 SLOESTRFSOLVELLKIICAFOLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYR 131	
Db		
QY	2 SLOESTRFSOLVELLKIICAFOLDTGLEYANSYNFAKKENNSPEHLKDEVSIIQSMGYR 61	
Db		
QY	132 NRAKRLLOSPEPNSLQETSLVSQSLNLTGTVRTLRTKRIQIOPKTSYVIELGSDSSEDTV 191	
Db		
QY	62 NRAKRLLOSPEPNSL-ETSLVSQSLNLTGTVRTLRTKRIQIOPKTSYVIELGSDSSEDTV 120	
Db		
QY	192 NKATYCSVGDOELLQITPQGTREISLDSAKKAACFESETDVTNTEHHQPSNNDLNTEK 251	
Db		
QY	121 NKATYCSVGDOELLQITPQGTREISLDSAKKAACFESETDVTNTEHHQPSNNDLNTEK 180	
Db		
QY	252 RAARHPEKYQGSVSNLHVPCGNTNTHASSLQHENSSLLLTKDRMNVEKAFCNKSQKP 311	
Db		
QY	181 RAAERHPEKYQGSVSNLHVPCGNTNTHASSLQHENSSLLLTKDRMNVEKAFCNKSQKP 240	
Db		
QY	312 GLARSOHNRWAGSKETCNDRTPTSTKTKVDLNADPLCERKEWNKQKLPCSENPRDTEVP 371	
Db		
QY	241 GLARSOHNRWAGSKETCNDRTPTSTKTKVDLNADPLCERKEWNKQKLPCSENPRDTEVP 300	
Db		
QY	372 WITLNSIOKVNWFSRSDLLGSDSDHGESESNKAVADVLVDLNEVDYSGSSEKIDL 431	
Db		
QY	301 WITLNSIOKVNWFSRSDLLGSDSDHGESESNKAVADVLVDLNEVDYSGSSEKIDL 360	
Db		
QY	432 LASDPHEALICKSERVHKSVESNIEDKIFGKTYRKASLPNLNLSHVTENLIIGAFVTEPQ 491	
Db		
QY	361 LASDPHEALICKSERVHKSVESNIEDKIFGKTYRKASLPNLNLSHVTENLIIGAFVTEPQ 420	
Db		
QY	492 IIOERPLTNLKKRRPTSLGHPEDIKKADLAVQKTPENINGTNOTQONGQVMNITNS 551	
Db		
QY	421 IIOERPLTNLKKRRPTSLGHPEDIKKADLAVQKTPENINGTNOTQONGQVMNITNS 480	
Db		
QY	552 GHENKTKGDSIQNEKNPNPTESLEKESAFKTKAEPITSSSISNMELELNHNSKAPKNRL 611	
Db		
QY	481 GHENKTKGDSIQNEKNPNPTESLEKESAFKTKAEPITSSSISNMELELNHNSKAPKNRL 540	
Db		
QY	612 RRKSSTRHITHALELVYSRNLSPNCTELQIDSCSSSEETKKKYKNQMPVRHSRNLQIMEG 671	
Db		
QY	541 RRKSSTRHITHALELVYSRNLSPNCTELQIDSCSSSEETKKKYKNQMPVRHSRNLQIMEG 600	
Db		
QY	672 KEPATGAKSKNKNPEOTSKRHSDSTPELKLTNAPGSFTKCSNTSELKEFVNPSLPREK 731	
Db		

Db	601 KEPATGAKSKNKNPEOTSKRHSDSTPELKLTNAPGSFTKCSNTSELKEFVNPSLPREK 660	
QY	732 EEKLETVKVSNNAEPRKDLMLSGERVLQTERSVESSISILVPGCTDYGTQBSISLLEVSTL 791	
Db		
QY	661 EEKLETVKVSNNAEPRKDLMLSGERVLQTERSVESSISILVPGCTDYGTQBSISLLEVSTL 720	
Db		
QY	792 GKAKTEPNKCVSOCAAPENPKGLIHGCSKDNNDTGEFKYPLGHEVNHNSRETSEMESE 851	
Db		
QY	721 GKAKTEPNKCVSOCAAPENPKGLIHGCSKDNNDTGEFKYPLGHEVNHNSRETSEMESE 780	
Db		
QY	852 LDAQYLQNTFKVSKROSFAFALFNPNAEPCATFSAHSGSLKKQSPKVTFCQSKENOG 911	
Db		
QY	781 LDAQYLQNTFKVSKROSFAFALFNPNAEPCATFSAHSGSLKKQSPKVTFCQSKENOG 840	
Db		
QY	912 KNESNIKPVQTVNITAGFPVVGQKDPVDNAKCSIKGGSREFCLSSQFRGNETGLITPNKH 971	
Db		
QY	841 KNESNIKPVQTVNITAGFPVVGQKDPVDNAKCSIKGGSREFCLSSQFRGNETGLITPNKH 900	
Db		
QY	972 GLLQNPYRIPPLPIKSFVKTKCKNLLLENFEEHSMSPEREMGNENIPSTVSTISRNNI 1031	
Db		
QY	901 GLLQNPYRIPPLPIKSFVKTKCKNLLLENFEEHSMSPEREMGNENIPSTVSTISRNNI 960	
Db		
QY	1032 RENNVKEASSNNINEVCSSSTNEVSSSINEIGSSDENIQAELGNRRPKLNAMRLGVLP 1091	
Db		
QY	961 RENNVKEASSNNINEVCSSSTNEVSSSINEIGSSDENIQAELGNRRPKLNAMRLGVLP 1020	
Db		
QY	1092 EYVKQSLPGSNCKHPEIKKOEYEEVQTVNTDFSPYLIISDNLEQPMGSSHASQVSETPD 1151	
Db		
QY	1021 EYVKQSLPGSNCKHPEIKKOEYEEVQTVNTDFSPYLIISDNLEQPMGSSHASQVSETPD 1080	
Db		
QY	1152 DLLDDGEIKEDTFAENDIKESSAVFSKSVOKGELSRSPFTHTHLAOGYRRGAKKLES 1211	
Db		
QY	1081 DLLDDGEIKEDTFAENDIKESSAVFSKSVOKGELSRSPFTHTHLAOGYRRGAKKLES 1140	
Db		
QY	1212 SEENLSSEDEELPCFOHLLPGKVNNIPQSOTRUSTVATECLSKNTENLLSLKNSLDCS 1271	
Db		
QY	1141 SEENLSSEDEELPCFOHLLPGKVNNIPQSOTRUSTVATECLSKNTENLLSLKNSLDCS 1200	
Db		
QY	1272 NOVILAKASQEHLSSETKCSASLFSQCSSELEDLTANTNTQDPFLIGSSKQMRHOSQ 1331	
Db		
QY	1201 NOVILAKASQEHLSSETKCSASLFSQCSSELEDLTANTNTQDPFLIGSSKQMRHOSQ 1260	
Db		
QY	1332 GVLGSKELVSDDEERGTLGEENNBEOQSMDSNLGAAAGCESETSVSEDCGLSSQSDI 1391	
Db		
QY	1261 GVLGSKELVSDDEERGTLGEENNBEOQSMDSNLGAAAGCESETSVSEDCGLSSQSDI 1320	
Db		
QY	1392 LTTQQRDTQWQHNLIKLOQEMAELEAVLEQHGSOPIINSYPSIISDSSALEDLRNPQOSTSE 1451	
Db		
QY	1321 LTTQQRDTQWQHNLIKLOQEMAELEAVLEQHGSOPIINSYPSIISDSSALEDLRNPQOSTSE 1380	
Db		
QY	1452 KAVLTISQKSEYPISONPEGLSADKFEVSADSTSKNKEPGVERSSPSKCPSLDDRWMH 1511	
Db		
QY	1381 KAVLTISQKSEYPISONPEGLSADKFEVSADSTSKNKEPGVERSSPSKCPSLDDRWMH 1440	
Db		
QY	1512 SCGSLQNRNYPQSEBELIKVVDVEEQOLBESGPHDITETSYLPRQDLEGTPLYESGISLF 1571	
Db		
QY	1441 SCGSLQNRNYPQSEBELIKVVDVEEQOLBESGPHDITETSYLPRQDLEGTPLYESGISLF 1500	
Db		
QY	1572 SDDPESDPSEDRAPEARSARVGNIPSSTSALKVPOLKVAESNAQSPAAHTTDTAGYNAMEES 1631	
Db		
QY	1501 SDDPESDPSEDRAPEARSARVGNIPSSTSALKVPOLKVAESNAQSPAAHTTDTAGYNAMEES 1560	
Db		
QY	1632 VSREKPELTASTERVNKRKSMVYSGLTPEEFMLVYKFAKKHHTLTNLNLTETEETHVVMKT 1691	
Db		
QY	1561 VSREKPELTASTERVNKRKSMVYSGLTPEEFMLVYKFAKKHHTLTNLNLTETEETHVVMKT 1620	
Db		
QY	1692 DAEFVVCERTLKYLELGIAGGKWWYSYFWVTQSIKERKMLNEHDFEVRGDVVVNGRHHQPKR 1751	
Db		
QY	1621 DAEFVVCERTLKYLELGIAGGKWWYSYFWVTQSIKERKMLNEHDFEVRGDVVVNGRHHQPKR 1680	
Db		
QY	1752 ARESQDRKIFRGLEICCYGPTNMPDQLEWVQLCASVVKELSSFTLGTGVHPVVVQ 1811	
Db		
QY	1681 ARESQDRKIFRGLEICCYGPTNMPDQLEWVQLCASVVKELSSFTLGTGVHPVVVQ 1740	
Db		

```
QY 1812 PDATWEDNGFHAIGOMCEAPVVTREWLDSVALYOCQELDTYLIPOIPHSY 1863
|||||
Db 1741 PDATWEDNGFHAIGOMCEAPVVTREWLDSVALYOCQELDTYLIPOIPHSY 1792

RESULT 3
O46485 PRELIMINARY; PRT; 1141 AA.
AC O46485;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer
DE susceptibility protein) (Fragment).
GN BRCA1.
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX NCBI_TaxID=9595;
RN [1]
RP SEQUENCE FROM N.A. PubMed=9462745;
RX MEDLINE=98122577;
RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,
RA Fodor S.P.A., Brody L.C., Collins F.S.;
RT "Evolutionary sequence comparisons using high-density oligonucleotide
RT arrays."
RL Nat. Genet. 18:155-158(1998).
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE
CC IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC EMBL; AF019076; AAC39584.1; -.
DR InterPro; IPR002378; Brst.CancerI.
DR PRINTS; PR00493; BRSTCANCER1.
KW DNA-binding; Nuclear protein; Anti-oncogene.
FT NON_TER 1
FT DOMAIN 277 284 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 427 430 POLY-LYS.
FT DOMAIN 985 1006 LEUCINE-ZIPPER (POTENTIAL).
FT NON_TER 1141 1141
SQ SEQUENCE 1141 AA; 127339 MW; A09531803BB99C3D CRC64;

Query Match 60.1%; Score 5795; DB 6; Length 1141;
Best Local Similarity 98.1%; Pred. No. 3.8e-291;
Matches 1119; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

QY 225 ACEFSETDVTNTEHQPSNNDLNTTEKRAAERHPEKYGCVSSVNLHVPCGTNTTHASSLQ 284
|||||
Db 1 ACEFSERDVTNTEHQPSNNDLNTTEKRAERHPEKYGCVSSVNLHVPCGTNTTHASSLQ 60

QY 285 HENSLLLTKDRMNYEKAFCNKKOPGLARSOHNRWAGSKETCNDRTPTTEKKVDLNA 344
|||||
Db 61 HENSLLLTKDRMNYEKAFCNKKOPGLARSOHNRWAGSKETCNDRTPTTEKKVDLNA 120

QY 345 DPLCERKENNKOKLPCSENPRDTEDPVITLNSISQKYNWFSDDELGGSDSDHGESE 404
|||||
Db 121 DPLCERNENWKOKLPCSENPRDTEDPVITLNSISQKYNWFSDDELGGSDSDHGESE 180

QY 405 SNAKVADVLVNLVEYDEYSGSEKIDLLASDPHEALICKSERVSKSVESNIEDKIFGKT 464
|||||
Db 181 SNAKVADVLVNLVEYDEYSGSEKIDLLASDPHEALICKSERVSKSVESNIEDKIFGKT 240

QY 465 YRKASLPNLSHVNTENLIIGAFVTPQIIQERPLTNLKKRRRPTSLGHPEDFIKKADLA 524
|||||
Db 241 YRKASLPNLSHVNTENLIIGAFVTPQIIQERPLTNLKKRRRATSLGHPEDFIKKADLA 300

QY 525 VQKTPMINQGTNQEQGVNMINTNSGHENKTKGDSIQNEKNPNPTESLEKESAFKTKA 584
|||||
Db 301 VQKTPMINQGTNQEQGVNMINTNSGHENKTKGDSIQNEKNPNPTESLEKESAFKTKA 360

QY 585 EPISSISNMELELNHNKAPKNRLRRKSTRIHALELVVSRNLSPPNCTELQIDSC 644
|||||
Db 361 EPISSISNMELELNHNKAPKNRLRRKSTRIHALELVVSRNLSPPNCTELQIDSC 420
```

```
QY 645 SSSBEIKKKYNQMPVHRHSRLQLMEGKEPATGAKKSNKPNEBOTSKRHSDSTFPELKLTN 704
|||||
Db 421 SSSBEIKKKYNQMPVHRHSRLQLMEGKEPATGAKKSNKPNEBOTSKRHSDSTFPELKLTN 480

QY 705 APGSFTKCSNTSELKEFVNPSLPREEKEEKLTVKVSNNAEADPKDMLSGERVLQTSRV 764
|||||
Db 481 APGSFTNCSNTSELKEFVNPSLPREEKEEKLTVKVSNNAEADPKDMLSGERVLQTSRV 540

QY 765 ESSISLVPGTDTYGTQBSISLLEVTSLGAKTEPNKCVSQCAAFENPKGLIHGCSKDNRN 824
|||||
Db 541 ESSISLVPGTDTYGTQBSISLLEVTSLGAKTEPNKCVSQCAAFENPKGLIHGCSKDTN 600

QY 825 DTEGFKYPLGHEVNHRSRTSITEMESELDAOYLQNTFKVSKRSQSPALSPNCAEEECAT 884
|||||
Db 601 DTEGFKYPLGHEVNHRSRTSITEMESELDAOYLQNTFKVSKRSQSPALSPNCAEEECAT 560

QY 885 FSAHSGSLKKOSPKVTFECEQKEENQGNKESNIRPVQTVNITAGFPVVGQKDKPVDNAKC 944
|||||
Db 661 FSAHRSRLKKOSPKVTFECEQKEENQGNKESNIRPVQTVNITAGFPVVGQKDKPVDYAKC 720

QY 945 SIKGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSFVTKCKKNLLEENFE 1004
|||||
Db 721 SIKGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSFVTKCKKNLLEENFE 780

QY 1005 EHSMSPEREMGNENIPSTVSTISRNNIRENVFKEAASSNINEVSGSSTNEVSGSINEIGSS 1064
|||||
Db 781 EHSMSPEREMGNENIPSTVSTISRNNIRENVFKEAASSNINEVSGSSTNEVSGSINEIGSS 840

QY 1065 DENTQAEALGRNRGPKLNAMRLGVLPQEVYKOSLPGSNCKHPEIKKQEEYEVVQVNTDF 1124
|||||
Db 841 DENTQAEALGRNRGPKLNAMRLGVLPQEVYKOSLPGSNCKHPEIKKQEEYEVVQVNTDF 900

QY 1125 SPYLISDNLEQPMGSSHASQVCSPTDLDLDDGETKEDTSAENDIKESSAVFSKVQKG 1184
|||||
Db 901 SPYLISDNLEQPMGSSHASQVCSPTDLDLDDGETKEDTSAENDIKESSAVFSKVQKG 960

QY 1185 ELSRSPSPFTHTLAQGYRRGAKKLESSEENLSSEDELPCHLFGKVNINPSQSTRH 1244
|||||
Db 961 ELSRSPSPFTHTLAQGYRRGAKKLESSEENLSSEDELPCHLFGKVNINPSQSTRH 1020

QY 1245 STVATECLSKNTEENLSKNSLNDSCSNQVILAKASQEHLSSEETKCSASFSSQCCSELE 1304
|||||
Db 1021 STVATECLSKNTEENLSKNSLNDSCSNQVILAKASQEHLSSEETKCSASFSSQCCSELE 1080

QY 1305 DLTANTNTQDFPLTGSSQMRHOSQCVGLSDKELVSDDEERGTLGLENNQEEQSDMSN 1364
|||||
Db 1081 DLTANTNTQDFPLTGSSQMRHOSQCVGLSDKELVSDDEERGTLGLENNQEEQSDMSN 1140

QY 1365 L 1365
Db 1141 L 1141

RESULT 4
O46484 PRELIMINARY; PRT; 1141 AA.
AC O46484;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer
DE susceptibility protein) (Fragment).
GN BRCA1.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98122577; PubMed=9462745;
RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,
RA Fodor S.P.A., Brody L.C., Collins F.S.;
```

"Evolutionary sequence comparisons using high-density oligonucleotide arrays.";
RL Nat. Genet. 18:155-158(1998).
CC -I- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE
CC IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
DR EMBL; AF019075; AAC39583.1; -.
DR InterPro; IPR002378; Brst_cancer1.
DR PRINTS; PR00493; BRSTCANCER1.
KW DNA-binding; Nuclear protein; Anti-oncogene.
FT NON_TER 1 1
FT DOMAIN 277 284 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 427 430 POLY-LYS.
FT DOMAIN 505 509 POLY-GLU.
FT DOMAIN 985 1006 LEUCINE-ZIPPER (POTENTIAL).
FT NON_TER 1141 1141
SQ SEQUENCE 1141 AA; 127399 MW; D72660B3CA6632D5 CRC64;

Query Match 59.9%; Score 5776; DB 6; Length 1141;
Best Local Similarity 97.9%; Pred. No. 3.6e-290;
Matches 1117; Conservative 6; Mismatches 18; Indels 0; Gaps 0;

QY 225 ACFFSETDVTNTEHHOPSNNDLNTTEKRAAERHPEKYQGVSSVNLHVEPCGTNTHASSLQ 284
Db 1 ACFFSETDVTNTEHHOPSNNDLNTTEKRAERHPEKYQGVSSVNLHVEPCGTNTHASSLQ 60

QY 285 HENSLLLTKDRMNVKAEFCNKSQPLARSQHNWAGSKETCNDRTPTSTKPKVDLNA 344
Db 61 HENSLLLTKDRMNVKAEFCNKSQPLARSQHNWAGSKETCNDRTPTSTKPKVDLNA 120

QY 345 DPLCERKWNKQLPCSENPRDTEVPWITLNSSIQVNEFWSRSDLLGSDSDSHDGESE 404
Db 121 DPLCERKWNKQLPCSENPRDTEVPWITLNSSIQVNEFWSRSDLLGSDSDSHDGESE 180

QY 405 SNAKVADVLVDNEVDYSSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKT 464
Db 181 SNAKVADVLVDNEVDYSSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKT 240

QY 465 YRKASLPNLSHVTENLIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPDEFIKRADLA 524
Db 241 YRKASLPNLSHVTENLIIGAFVTEPQIIQERPLTNKLRKRATSGLHPDEFIKRADLA 300

QY 525 VQKTPWINGOTNOTONGOVNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKA 584
Db 301 VQKTPWINGOTNOMEQGVNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKA 360

QY 585 EPISSSISNMELELNTHNSKAPKNRLRRKSTRHIALELVVSRNLSPPNCTELQIDSC 644
Db 361 EPISSSISNMELELNTHNSKAPKNRLRRKSTRHIALELVVSRNLSPPNCTELQIDSC 420

QY 645 SSSEEIKKKYNQMPVHRHSNRLQMEGKEPATGAKKSNKPNQTSKRHSDTPELKLTN 704
Db 421 SSSEEIKKKYNQMPVHRHSNRLQMEGKEPATGVKKSNEQTSKRHSDTPELKLTN 480

QY 705 APOGFTKCSNTSELKEFVNPSLPREKEEKLTVKYVSNNAEDPKDMLSGSERVLQTSRV 764
Db 481 APOGFTKCSNTSELKEFVNPSLPREKEEKLTVKYVSNNAEDPKDMLSGSERVLQTSRV 540

QY 765 ESSISIVPGTDTGTOESISLLEVTSLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDRN 824
Db 541 ESSISIVPGTDTGTOESISLLEVTSLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDRN 600

QY 825 DTGFKYPLGHEVNHRSRETSIEMEESLDAQYLQNTFKVSKROSFALFSPNGNAEECAT 884
Db 601 DTGFKYPLGHEVNHRSRETSIEMEESLDAQYLQNTFKVSKROSFALFSPNGNPRECAT 660

QY 885 FSAHSGSLKQSPKVFTECEQKEENQKNSNIKPQVTNITAGFPVQOKPKVDNAKC 944
Db 661 FSAHSGSLKQSPKVFTECEQKEENQKNSNIKPQVTNITAGFPVQOKPKVDNAKC 720

QY 945 SIKGSRFCFLSSQFRNETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKKNLLENPE 1004
Db 721 SIKGSRFCFLSSQFRNETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKKNLLENPE 780

QY 1005 EHSMSPEREMGNENIPSTVSTISRNNIRENVFKEASSNNINEVGSSTNEVGSSINEIGSS 1064
Db 781 EHSMSPEREMGNENIPSTVSTISRNNIRENVFKEASSNNINEVGSSTNEVGSSINEIGSS 840

QY 1065 DENIQAEELGRNRPKLNAMRLGLVQPEVYKQSLPSNCKHPKQIEYEEVQVTNTDF 1124
Db 841 DENIQAEELGRNRPKLNAMRLGLVQPEVYKQSLPSNCKHPKQIEYEEVQVTNTDF 900

QY 1125 SPYLISDNLQPMGSSHASOVCSPTDDLLDDGEIKEDTSFAENDIKESSAVFSKSVQKG 1184
Db 901 SPCLISDNLQPMGSSHASOVCSPTDDLLDDGEIKEDTSFAENDIKESSAVFSKSVQKG 960

QY 1185 ELSRSPSPFTHTHLAQYGRGAKKLESSEENLSEDEELPCFOHLLFGKVNIPQSSTRH 1244
Db 961 ELSRSPSPFTHTHLAQYGRGAKKLESSEENLSEDEELPCFOHLLFGKVNIPQSSTRH 1020

QY 1245 STVATECLSKNTEENLLSKNSLNDSCNQVILAKASQEHHLSEETKCSASLFSQCSSELE 1304
Db 1021 STVATECLSKNTEENLLSKNSLNDSCNQVILAKASQEHHLSEETKCSASLFSQCSSELE 1080

QY 1305 DLTANTNTODPFLIGSSKOMRHOSESQGVGLSDKELVSDDEERGTLGLEENNOEQSMDSN 1364
Db 1081 DLTANTNTODPFLIGSSKOMRHOSESQGVGLSDKELVSDDEERGTLGLEENNOEQSMDSN 1140

QY 1365 L 1365
Db 1141 L 1141

RESULT 5
O46486 PRELIMINARY; PRT; 1141 AA.
AC O46486;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer
DE susceptibility protein) (Fragment).
GN BRCA1.
OS Pongo. pygmaeus (Oranoutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX NCBI_Taxid=9600;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98122577; PubMed=9462745;
RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,
Fodor S.P.A., Brody L.C., Collins F.S.;
RT "Evolutionary sequence comparisons using high-density oligonucleotide
RT arrays.";
RL Nat. Genet. 18:155-158(1998).
CC -I- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE
CC IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
DR EMBL; AF019077; AAC39585.1; -.
DR InterPro; IPR002378; Brst_cancer1.
DR PRINTS; PR00493; BRSTCANCER1.
KW DNA-binding; Nuclear protein; Anti-oncogene.
FT NON_TER 1 1
FT DOMAIN 277 284 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 427 430 POLY-LYS.
FT DOMAIN 985 1006 LEUCINE-ZIPPER (POTENTIAL).
FT NON_TER 1141 1141
SQ SEQUENCE 1141 AA; 127450 MW; A4C692E5779091F1 CRC64;

Query Match 58.8%; Score 5671; DB 6; Length 1141;
Best Local Similarity 96.3%; Pred. No. 9.6e-285;
Matches 1099; Conservative 11; Mismatches 31; Indels 0; Gaps 0;

QY 225 ACFFSETDVTNTEHHOPSNNDLNTTEKRAAERHPEKYQGVSSVNLHVEPCGTNTHASSLQ 284
Db 1 ACFFSETDVTNTEHHOPSNNDLNTTEKRAERHPEKYQGVSSVNLHVEPCGTNTHASSLQ 60

QY 285 HENSSLLLTQDRMVEKAFCKNSKOPGLARSOHNRWAGSKETCNDRTPTSTTEKKVVDLNA 344
DB 61 HENSSLLLTQDRMVEKAFCKNSKOPGLARSOHNRWAGSKETCNDRTPTSTTEKKVVDLNA 120
QY 345 DPLCKERKNKQKLPCCSENPRDTEVPWITLNSSTQKVNEFWSRDELLGSDSDHGGSE 404
DB 121 DPLCKERKNKQKLPCCSENPRDTEVPWITLNSSTQKVNEFWSRDELLGSDSDHGGSE 180
QY 405 SNAKVADVLVDVNEVDEYSGSEKIDLLASDPHEALICKSERVHKSVESNTEDEKIFGKT 464
DB 181 SNAKVADVLVDVNEVDEYSGSEKIDLLASDPHEALICKSERVHKSVESNTEDEKIFGKT 240
QY 465 YRKASLPNLSHVTEENLIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPDEFIKKADLA 524
DB 241 YRKASLPNLSHVTEENLIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPDEFIKKADLA 300
QY 525 VQKTPENINGTNOTQONGOVNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKA 584
DB 301 VQKTPENINGTNOTQONGOVNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKA 360
QY 585 EPISSSISNWELELNTHNSKAPKNRLRKSSTRHIALELVVSRNLPSPNCTELQIDSC 644
DB 361 EPISSSISNWELELNTHNSKAPKNRLRKSSTRHIALELVVSRNLPSPNCTELQIDSC 420
QY 645 SSSEEEKKKYKNOMPVRHNRNLQMEGEPATCAKSKNKPNEOTSKRHSOTFFPELKLTN 704
DB 421 SSSEEEKKKYKNOMPVRHNRNLQMEGEPATCAKSKNKPNEOTSKRHSOTFFPELKLTN 480
QY 705 APGSFTKCSNTSELKEFVNPSLPREEKEKLETVKVSNNAPDKDMLSGERVLOTERSV 764
DB 481 APGSFTKCSNTSELKEFVNPSLPREEKEKLETVKVSNNAPDKDMLSGERVLOTERSV 540
QY 765 ESSISLVPGTDYGTQESTLSLEVTSLGKAKTEPNKCVSOCAAFENPKGLLHGCKSNRN 824
DB 541 ESSISLVPGTDYGTQESTLSLEVTSLGKAKTEPNKCVSOCAAFENPKGLLHGCKFDNRN 600
QY 825 DTGFKYPLGHEVNSRSTSIEMWSELDQAQYQNTFKVSKRQSFALPSNPGNAEECAT 884
DB 601 DTGFKYPLGHEVNSRSTSIEMWSELDQAQYQNTFKVSKRQSFALPSNPGNAEECAT 660
QY 885 FSAHSGSLKKQPKVTFECEQKEENOGKNESNIKPVQTNITAGFPVVGQDKPVDNAKC 944
DB 661 FSAHSGSLKKQPKVTFECEQKEENOGKNESNIKPVQTNITAGFPVVGQDKPVDNAKC 720
QY 945 SIKGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPKISPVKTKCKKNLLEENFE 1004
DB 721 SIKGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPKISPVKTKCKKNLLEENSE 780
QY 1005 EHSMSPEREMGNENIPSTVSTISRNIRENVFKEASSNINEVGSSTNEVGSSINEIGSS 1064
DB 781 EHSMSPEREMGNENIPSTVSTISRNIRENVFKEASSNINEVGSSTNEVGSSINEVGSS 840
QY 1065 DENIAELGRNRPKLNAMRLGVLOPEYIKOSLPGSNCKHPEIKKQBYEYVQTVNTDF 1124
DB 841 DENIAELGRNRPKLNAMRLGVLOPEYIKOSLPGSNCKHPEIKKQBYEYVQTVNTDF 900
QY 1125 SPYLISDNLEQPMGSSHASQVCEPDDLLDGEIKEDTSAENDIKESSAVFSSQVK 1184
DB 901 SPYLISDNLEQPMGSSHASQVCEPDDLLDGEIKEDTSAENDIKESSAVFSSQVK 960
QY 1185 ELSRSPSPPTHLAGQYRRGAKKLESSEENLSSDEDELPCFHLLFGKVNIPQSSTRH 1244
DB 961 ELSRSPSPPTHLAGQYRRGAKKLESSEENLSSDEDELPCFHLLFGKVNIPQSSTRH 1020
QY 1245 STVATECLSKNTEENLLSKNSLNDQSNQVILKASQEBHLSSEETKCSASLFSQCSSELE 1304
DB 1021 STVATECLSKNTEENLLSKNSLNDQSNQVILKASQEBHLSSEETKCSASLFSQCSSELE 1080
QY 1305 DLTANTNTODPELIGSSKOMRHQESQVGLSDKELVSDDEERGTLGLENNOEQSGSDSN 1364
DB 1081 DLTANTNTODPELIGSSKOMRHQESQVGLSDKELVSDDEERGTLGLENNOEQSGSDSN 1140

QY 1365 L 1365
DB 1141 L 1141
RESULT 6
O46487
ID O46487 PRELIMINARY; PRT: 1140 AA.
AC O46487; O28525;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer
DE susceptibility protein) (Fragment).
GN BRCA1.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98122577; PubMed=9462745;
RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,
RA Fodor S.P.A., Brody L.C., Collins F.S.;
RT "Evolutionary sequence comparisons using high-density oligonucleotide
RT arrays.";
RL Nat. Genet. 18:155-158(1998).
RN [2]
RP SEQUENCE OF 964-1028 FROM N.A.
RA Thompson M.E., Holt J.T.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE
CC -!- IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
DR EMBL; AF019078; AAC39586.1; -;
DR EMBL; U4730; AAB03212.1; -;
DR InterPro: IPR002378; Brst_cancer1.
DR PRINTS; PR00493; BRSTCANCER1.
KW DNA-binding; Nuclear protein; Anti-oncogene.
FT NON_TER 1
FT DOMAIN 276 283 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 984 1005 LEUCINE-ZIPPER (POTENTIAL).
FT CONFLICT 978 978 R -> Q (IN REF. 2).
FT NON_TER 1140 1140
SQ SEQUENCE 1140 AA; 126958 MW; EC1873CAFA5E32F7 CRC64;
Query Match 56.4%; Score 5443.5; DB 6; Length 1140;
Best Local Similarity 92.6%; Pred. No. 5.4e-273;
Matches 1057; Conservative 35; Mismatches 47; Indels 3; Gaps 3;
QY 225 ACEFSETDVTNTEHHQSPNSNDLNTTEKRAERHPKPYOGSSVSNLHVPCGNTTHASSIQ 284
DB 1 ACEFSEKIDITTEHHQSSNDLNTTEKHATERHPKPYOGSSVSNLHVPCGNTTHASSIQ 60
QY 285 HENSSLLLTQDRMVEKAFCKNSKOPGLARSOHNRWAGSKETCNDRTPTSTTEKKVVDLNA 344
DB 61 HEN-SLLLTQDRMVEKAFCKNSKOPGLARSOHNRWAGSKETCNDRTPTSTTEKKVVDLNA 119
QY 345 DPLCKERKNKQKLPCCSENPRDTEVPWITLNSSTQKVNEFWSRDELLGSDSDHGGSE 404
DB 120 NALYERKNWKKQKLPCCSENPRDTEVPWITLNSSTQKVNEFWSRDELLGSDSDHGGSE 179
QY 405 SNAKVADVLVDVNEVDEYSGSEKIDLLASDPHEALICKSERVHKSVESNTEDEKIFGKT 464
DB 180 SNAKVADVLVDVNEVDEYSGSEKIDLLASDPHEALICKSERVHKSVESNTEDEKIFGKT 239
QY 465 YRKASLPNLSHVTEENLIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPDEFIKKADLA 524
DB 240 YRKASLPNLSHVTEENLIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPDEFIKKADLA 299
QY 525 VQKTPENINGTNOTQONGOVNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKA 584
DB 301 VQKTPENINGTNOTQONGOVNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKA 584

Db	300	VQKTEIINOQTNOEQNGOVMMITNSAHENKTKGDSIQNEKNPNPIESLEESAFKTKA	359
QY	585	EPISSSISNMELELNHNSKAPKKNLRKSSSTRHHALELVVSRNLSPPNCTELQIDSC	644
Db	360	EPISSSINNELELNHNSKAPKKNLRKSSSTRHHALELVVSRNLSPPNCTELQIDSC	419
QY	645	SSSEETKKKKYNOMPVRHNSRNLMQMEGKEPATGAKSKNKPNEQTSKRHSDTPELKLTN	704
Db	420	SSSEETKKKKYNOMPVRHNSRNLMQMEGKESATGAKSKNKPNEQTSKRHASDTPELKLTK	479
QY	705	APGSFTKCSNTSELKEFVNPSLPREEKEKLETVKVSNNNAEDPKDMLSGERVLOTERS	764
Db	480	VPGSFTNCSTSE-KEFVNPSLPREEKEKLETVKVSNNNAEDPKDMLSGERVLOTERS	538
QY	765	ESSSISLVPDGTGQTESISLLEVSITLGAKTEPNKVCVSOCAAFENPKGLIHGCSKDNRN	824
Db	539	ESSSISLVPDGTGQTESISLLEVSITLGAKTERNKCMOCAAFAENPKELIHGCSSETRN	598
QY	825	DTGCFKYPGLGHEVNHNSRETSEIEMESELDAQYLQNTFKVSKROSFAFNSPGNAEECAT	884
Db	599	DTGCFKYPGLGSEVNHNSQETSIEIESESELDTQYLQNTFKVSKROSFAFNSPGNPEECAT	658
QY	885	FSAHSGSLKKQSPKVFECQKKEENCKNESNIKPVQTNITAGFPVVGOKDPVDNAKC	944
Db	659	FSAHSGSLKKQSPKVFECQKKEENCKQKOSNIKPVQTNITAGFVVCOKDPVDNAKC	718
QY	945	SIRGGSRFCLSSQFRNETGLITPNKHGILLQNPYRPPFPPIKSFVKTKCKNLLLENFE	1004
Db	719	SIRGGSRFCLSSQFRNETGLITPNKHGILLQNPYHIPPLFPVKSFVKTKCKNLLLENSE	778
QY	1005	EHSMSPEREMGNEN-IPSTVSTISRNIRENVPEKASSSINIEVGSSTNEVGSSINIEGS	1063
Db	779	EHSVSPERAVGNKNIIPSTVSTISHNNIRENAFKEASSSINIEVGSSTNEVGSSINIEVS	838
QY	1064	SDENIOAELGRNRPGLNMLRGVLQPEVYKOSLPCNSCKHPKIKKOEYEEVQTVNTD	1123
Db	839	SDENIOAELGRNRPGLNMLRGVLQPEVCOSLPTISNCKHPKIKKOEHEELVQTVNTD	898
QY	1124	FSPLYSDNLNLEQPMGSSHASQVSETPDLLDDGEIKEDTSFAENDIKESSAVFSKSQK	1183
Db	899	FSPLYSDNLNLEQPMGSSHASQVSETPDLLDDGEIKEDTSFAENDIKESSAVFSKSQK	958
QY	1184	GELSRSPSPFTHLAQGYRGAKKLESSEENLSDEDELPFCQHLHFGKVNIPSOSTR	1243
Db	959	GELSRSPSPFTHLAQGYRGAKKLESSEENLSDEDELPFCQHLHFGKVNIPSOSTR	1018
QY	1244	HSTVATECLSKNTEENLLSLKNSINDCSNOVILAKASQEHHLSEETKCSASFSSQCSL	1303
Db	1019	HSTVATECLSKNTEENLLSLKNSINDCSNOVILAKASQEHHLSEETKCSASFSSQCSL	1078
QY	1304	EDLTANTNTQDPFLIGSSKOMRHSQSGVGLSDKELVSDDERGTGLENNQDEQSMD	1363
Db	1079	EDLTANTNTQDPFLIGSSKOMRHSQSGVGLSDKELVSDDERGTGLENNQDEQSMD	1138
QY	1364	NL 1365	
Db	1139	NL 1140	
RESULT 7			
O46488			
ID	O46488	PRELIMINARY;	PRT; 1146 AA.
AC	O46488;		
DT	01-JUN-1998 (trEMBLrel. 06, Created)		
DT	01-JUN-1998 (trEMBLrel. 06, Last sequence update)		
DT	01-JUN-2001 (trEMBLrel. 17, Last annotation update)		
DE	Breast cancer type 1 susceptibility protein (Breast and ovarian cancer		
DE	susceptibility protein) (Fragment).		
GN	BRCA1.		
OS	Alouatta seniculus (Howler monkey).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Alouattinae;		
OC	Alouatta.		

OX	NCBI_TaxID=9503;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98122577; PubMed=9462745;		
RA	Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,		
RA	Fodor S.P.A., Brody L.C., Collins F.S.;		
RT	"Evolutionary sequence comparisons using high-density oligonucleotide		
RT	arrays.";		
RL	Nat. Genet. 18:155-158(1998).		
CC	-!- FUNCTION: NOT KNOWN. MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE		
CC	IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).		
CC	-!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).		
DR	EMBL; AF019079; AAC39587.1; -.		
DR	InterPro; IPR002378; BrstCancer1.		
DR	PRINTS; PR00493; BRSTCANCER1.		
KW	DNA-binding; Nuclear protein; Anti-oncogene.		
FT	NON_TER 1 1		
FT	DOMAIN 278 285		
FT	DOMAIN 200 203		
FT	DOMAIN 815 818		
FT	DOMAIN 1001 1004		
FT	DOMAIN 996 1005		
FT	NON_TER 1146 1146		
SQ	SEQUENCE 1146 AA; 128119 MW; BFD07EBBF851890E CRC64;		
Query Match 52.0%; Score 5014.5; DB 6; Length 1146;			
Best Local Similarity 86.2%; Pred. No. 7.9e-251;			
Matches 991; Conservative 55; Mismatches 90; Indels 13; Gaps 7;			
QY	226	CEPSETDVTNTEHHQPSNNDLNTTEKRAERHPEKYGGSSVSNLHVPCGTNTHASSLQH	285
Db	2	CEPSEKIDTNEYHQSSNNDLNTTEKHATEKHPEKQCQSSVSNLHVPCGTNTHASSLQH	61
QY	286	ENSSLLTLDKRMVKAERCNKSKQPLARSOHNRWAGSKETCNDRTSTEEKVLDNAD	345
Db	62	ENSSLLTLDKTLNKAERCNKSKQPLARSOHNRWAESEETCNDRTSTEEKVLDNAD	121
QY	346	PLCKEREMNKQKLPCCSENPR-DTEDEVPTITNSSIOKVNEWFSRDELGSDSDHGESE	404
Db	122	PLHGRKEMNKQKPPCCSENPRDDTEDIWIMLSSIOKVNEWFSRDELTSDDSHGGSE	181
QY	405	SNAKVADVLVNEVDEYSGSSEKIDLLASDPHEALICKSERVHSKVSNEIEDKIFGKT	464
Db	182	SNAKVAEALVLENEVDGYSSESSEKIDLLASDPHDHLICKSERVHCKSVSESSIEDKIFGKT	241
QY	465	YRKASLPLNLSHVTENLIIICAFVTEPQIIQERPLTNKLRKRPTSLGHPEDFIKKADLA	524
Db	242	YRKASLPLNLSHVTENLIIICAFVTEPQIIQERPLTNKLRKRPTSLGHPEDFIKKADLA	301
QY	525	VQKTEIINOQTNOEQNGOVMMITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKA	584
Db	302	VQKTEKINOQTNOEQNGOVMMITNSGHENKTKGDSIQNEKNPNPVSLEKES-FKSKA	360
QY	585	EPISSSISNMELELNHNSKAPKKNLRKSSSTRHHALELVVSRNLSPPNCTELQIDSC	644
Db	361	EPISSSISNMELELNHNSKASKNLRKSSSTRHHALELVVSRNLSPPNCTEVIDSC	420
QY	645	SSSEETKKKKYNOMPVRHNSRNLMQMEGKEPATGAKSKNKPNEQTSKRHSDTPELKLTN	704
Db	421	SSSEETKKKKYNOMPVRHNSRNLMQMEKERAARAKSKSPNEQTSKRHASDTPELKLTN	480
QY	705	APGSFTKCSNTSELKEFVNPSLPREEKEKLETVKVSNNNAEDPKDMLSGERVLOTERS	764
Db	481	IPGSFTNCSTNEFKEFVNPSLPREEKEKLETVKLSNNAKDPKDLMSGESVLOIERSV	540
QY	765	ESSSISLVPDGTGQTESISLLEVSITLGAKTEPNKVCVSOCAAFENPKGLIHGCSKDNRN	824
Db	541	ESSSILLIPDGTGQTESISLLEVSITLGAKTEPNKVCVSOCAAFENPKELIHGCSKSDTRN	600
QY	825	DTGCFKYPGLGHEVNHNSRETSEIEMESELDAQYLQNTFKVSKROSFAFNSPGNAEECAT	884
Db	601	GTEGLKYPGLPEVNYSOETSIDMRESELDTQYLQNTFKVSKROSFAFNSPGNPEKCAT	660

QY	885	FSHSGSLKKQSPKVTPECEQKEENOGKNESNIKPVQTVNITAGFPVVVGQDKPVDNAKC	944
DB	661	FSACSRSLKKQSPKVTPECEQKEENOGKESNIELVETVNTTAGFPMVCQDKPVDYARC	720
QY	945	SIKGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPKTSFVKTKCKKNLLEENFE	1004
DB	721	-IEGSRCLSSQFRGNETGLITPNKHGLLQNPYHMSPLIPTRSFVKTKCKKNLLEENSE	779
QY	1005	EHSMSPEREMGNEN-IPSTVSTISRNNTRENVEFKASSNNI-----NEVGSSTNEVGS	1056
DB	780	EHSMSPERAMGNKIIPSTVSTISINN-RENAFKETSSSYEVGSSTNEAGSSTNEVGS	838
QY	1057	SINEIGSSDENIQAEELGRNRGPKLAMLRLGLVQPEVYKSLPGSNCKHPKIKQYEEV	1116
DB	839	SINEVGSSENIQAEELGRNRGPKLAMLRLGLLQPEICKQSLPISDCKHPKIKQHEEV	898
QY	1117	VQTVNTDFSPYILSNLEQPMGSSHASQVCSSTPDDLLDDGEIKEKDTFAENDIKESSAV	1176
DB	899	VQTVNTDVSCLISYLNLEQHMGSSTSQVCSSTPDDNLLDDGEIKEKDTFAEYGIKETSTV	958
QY	1177	FKSVQKGELSPSPFTHTHLAQYRRGAKKLESSEENLSEDELPFOHLLFGKVVN	1236
DB	959	FKSVQVGGELSPSPFTHTHLAQYVRGAKKLESSEENLSEEEELPCFOHLLFGKVVN	1018
QY	1237	IPSQSTRHSTVATECLSKNTEENLLSKNSLNDSCNQVILAKASOEHLSEETKCSASLF	1296
DB	1019	TP-QSTRHSTSAECLSKNTEENLLSKSSLSDSCNQVILAKASQEHYLSSEETKCSGSLF	1077
QY	1297	SSQCSLEDLTANTWTDQDFLLGSSQKMRHQSOGVGLSKELVSDDEBRTGLERNQ	1356
DB	1078	SSQCSLEDLTNTWTDQDFLLDSSKMRROSESOGVGLSKELVSDDEBRTGLDEEDNQ	1137
QY	1357	EQQSDMSNL 1365	
DB	1138	EQQSDMSNL 1146	
RESULT 8			
ID	Q921D2	PRELIMINARY;	PRT; 1812 AA.
AC	Q921D2;		
DT	01-MAY-1999 (TremBLrel. 10, Created)		
DT	01-MAY-2002 (TremBLrel. 10, Last sequence update)		
DT	01-JUN-2002 (TremBLrel. 21, Last annotation update)		
DE	BRCA1.		
GN	BRCA1 OR BRCA1.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=129/SV;		
RX	MEDLINE=97026289; PubMed=8872468;		
RA	Szabo C.I., Wagner L.A., Francisco L.V., Roach J.C., Argonza R.,		
RA	King M.C., Ostrander E.A.;		
RT	"Human, canine and murine BRCA1 genes: sequence comparison among		
RT	species."		
RL	Hum. Mol. Genet. 5:1289-1298(1996).		
CC	-!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.		
DR	EMBL; U68174; AAD00168.1; -		
DR	MGI; MGI:104537; Brca1.		
DR	InterPro; IPR001357; Brct.		
DR	InterPro; IPR002378; Brst_cancerI.		
DR	InterPro; IPR001547; GH_5.		
DR	InterPro; IPR001841; Znf_ring.		
DR	Pfam; PF00533; BRCT; 2.		
DR	Pfam; PF00097; zf-C3HC4; 1.		
DR	PRINTS; PR00493; BRSTCANCER1.		
DR	SMART; SM00292; BRCT; 2.		
DR	SMART; SM00184; RING; 1.		
DR	PROSITE; PS00172; BRCT; 2.		
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.		

DR	PROSITE; PS00518; 2F_RING_1; 1.		
KW	Zinc-finger.		
SQ	SEQUENCE 1812 AA; 198803 MW; 1A61B60230365B63 CRC64;		
Query Match 50.6%; Score 4885; DB 11; Length 1812;			
Best Local Similarity 56.0%; Pred. No. 7.2e-244;			
Matches 104; Conservative 265; Mismatches 485; Indels 70; Gaps 32;			
QY	1	MDLSALRVEEVQNVINAMQKILECPICILELKEPYSTKCDHIFCFKMLKLLNKKGPSSQ	60
DB	1	MDLSAVQIQEVNVLHMQKILECPICILELKEPYSTKCDHIFCFKMLKLLNKKGPSSQ	60
QY	61	CPLCKNDITKRSLOESTRFSQVLEELLKIIICAFQDLDGLEYANSYNFAKKENNSPEHLKD	120
DB	61	CPLCKNEITKRSLOGSTRFSQVLEELLKIIICAFQDLDGLEYANSYNFAKKENNSPEHLKD	120
QY	121	EVSIIQSGYNNRKRRLQSEPNPSLOETSLVSQNLGTVRLTKRQIQRQKTSVYI	180
DB	121	EASIIQSGYNNRKRRLQSEPNPSLOETSLVSQNLGTVRLTKRQIQRQKTSVYI	180
QY	181	ELGSDSSEDTVNKATYCSVGOELLQITPQGTDRDEISLSAKKAAACERSETDVTTEHHQ	240
DB	180	ELGSDSSEDTVNKATYCSVGOELLQITPQGTDRDEISLSAKKAAACERSETDVTTEHHQ	240
QY	241	PSNNDLNTTKRAAERHPEKYQGSVSNLHVPCGTNTHASLQHENSSLTLDKRMNVE	300
DB	239	CS-DLNPNTENHATERHPEKQCSISINVCVPCGTDAHASSLOPETSLLLIEDRNAE	297
QY	301	KAFCNKSQKQGLARSHNRWAGSKETCNDRTPTSTKRYDLNADPLCERKWNKQKLP	360
DB	298	KAFCNKSQKQGLARSHNRWAGSKETCNDRTPTSTKRYDLNADPLCERKWNKQKLP	360
QY	361	SENPRDTVPWITLNSIOKVNWFSDGLGSDSDHGESESNKAVADVLDVNEVD	420
DB	358	PENSGATDVPWITLNSVSVKVNWFSGTGMELTSDSASARRHESNAEAAVVEVNEVD	417
QY	421	EYSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKYRKKASPLNLSHVTEN	480
DB	418	GGFSSRRKTLVTPDHPHTLACKSGRDFSKPVEDNISDKIFKSVQRKSGRPHLNHVT	476
QY	481	LIIICAFVTEPOIIQERPLTNKLRKRPSTGLHPEDFTKKADLA-VQKTPMINOQTQT	539
DB	477	-IITFTITEPOITQEQPFTNKLKRKR--STSLQPEDFIKKADSAQVQRTPDINOQTDLM	533
QY	540	BQNGVWNITNSGHENKTKGDSIONENPNPTESLESKAEKTAEPISSTSNNELELN	599
DB	534	EPNQAVSTTSNQENQJAGSNLQKESAPHTESLRKEPASTAGAKSISNSVSDLEVLN	593
QY	600	IHNSKAPKNNLRKSSPTRHITHALELVVSRNLSPNCTELQIDSCSSSEETKKKYNNQMP	659
DB	594	VHSSKAPKNNLRKSSPTRHITHALELVVSRNLSPNCTELQIDSCSSSEETKKNHNNQMP	652
QY	660	VHRSNLQMEGKEPATGAKKSNKPNQETSKRHSDTTPPELKLTNAGPSFTKCSNTSELK	719
DB	653	AGHLREPLQIEDTEPAADAKK-NEPNEHIRKRASDAPPEEKLMKAGLLTSCSPRSQ	711
QY	720	EFVNPSLPREEKEKLETVKVSNAEDPKDLMLSGE-RVLOTERSVSSSTSLVPGTDVG	778
DB	712	GPVNPSP-PORTGTQLETRQMSDAKELGDRVLGGEPSGKTTDRSEESTSLVPTDID	770
QY	779	TQESISLLEVTGLGAKTEPNKVCYQCAAFENPKGLIHGCSKDNNDNTEGPKYPLGHEVN	838
DB	771	TQNSVSLDAHTVRYARTGSAQCMQTFVASENPKELVHG-SNNAGSGTEGLKPLRLHALN	829
QY	839	HSRTSTSEMESELDQYLNQTFKVKRQSFALSPNPGNAEEECATFASHGSLKKQSPK	898
DB	830	LSQE-KVEMEDSELDQYLNQTFKVKRQSFALSPNPGNAEEECATFASHGSLKKQSPK	884
QY	899	VTFPECKEENOGKNESNIKPVQTVNITAGFPVVVGQDKPVDNAKCSIKGGSFCLSSQF	958
DB	885	VTAKGKOR-ERQGOEEFEISHVQVAATVGLPVGEGKLAADTWCD--RGSRLCPSSH	941
QY	959	RGNETGLITPNKHGLLQNPYRIPPLFPKTSFVKTKCKKNLLEENFEHSMSPEREMGNEN	1018

```
Db 942 RSGNGLSATGKSGISONSFHFKQSVSPIRSISIKTDNKPLETEGFERHRTSTSEMAVGNEN 1001
QY 1019 I-PSTVSTISRNNTRENVFWEKASSNNINEVGSSSTNEVGSSDENEIQAEPLGNRG 1077
Db 1002 ILQSTVHTVSLNN-RGNACQAGS-----GSIHEVCSTGDSFGPQGLGRNRG 1046
QY 1078 PKLNAMLRLGLVLOPEVYKQSLPGSNCKHPEIKKQYEEVVQTVNTDFSPYLDISNLBQPM 1137
Db 1047 PKVNTVPPLDSMQPGVCQQSPVSD-KYLEIKKQGEAVC----ADFSPLCFSDHLQSQM 1101
QY 1138 GSSHASQVCSETPDDLLDDGEIKEDTFAENDIKESSAVFSKSVQKGLSRSPSPFTHTH 1197
Db 1102 -SGKVFQVCSETPDDLLDDVEIOGHTSFGEGDIMERSAVFGSILRRRESSRSPSPVTHAS 1160
QY 1198 LAQGYRRGAKLLESSRENLSSEDELPFOHLLFGKVNNTIPSQSTRHSTVATCLSKNTE 1257
Db 1161 KSQLHRASRKLESSESDSTEDDLPCFQHLL-SRLSNTP-ELTRCSAVTOGIPKAE 1218
QY 1258 ENLLSKNSLNDCSNQVILAKASQEHLSSETRKCSALFSQCSSELDLTANTNTQDPFL 1317
Db 1219 GTQAPWKSSSDCNEVIMTEASQEHQFSEDPRCSGRMFSSQNSAAQGSTANANSQDSNF 1278
QY 1318 IGSSKQMRHOSQGVGLSKELVSDDEERGTLGLENNQEBQSDMSNLGA-ASGCSEST 1376
Db 1279 IPPSKQRSHQCGNEAEFLSKELISDNEEMATCLEEDNDQEE--DSIIPDSEASGYESET 1336
QY 1377 SVSDCSGLSSQSDILTTQORDTMOHLNLIKLOQEMAELEAVLEOHGSGPSNSYPSIISDS 1436
Db 1337 NLSEDC-----SQSDILTTQORDTMOHLNLIKLOQEMAELEAVLEOHGSGPSNSYPSIISDS 1392
QY 1437 SALEDLRNPQSTSEKAVLTQSOKSSEYPISONPEGLSADKFEVS-ADSSTSKNKEPGVER 1495
Db 1393 CALEDLPDLEPNMGAAIILTSKNINENPVSONLKSACDDKRFQLQHLGEGPTSGDDESGMR 1452
QY 1496 SSPSKCPSLDRWYMHSCSGSLQNRNYPQSEBELIKVYVDVEEQLEESGPHDLTETSYLPR 1555
Db 1453 PSPPKSPLAGSGSAHCSRHLOKRNPSQFELLQAPGSE-----ASSEPHNSTGQSCCLPR 1508
QY 1556 QDLEGTPLYESGLSIFSD-DPESDPSDRAPESARVGNIPSTISALKVPOLKVAESAQSP 1614
Db 1509 RELEGTPLYESGLSIFSSRDPESEPK-----EPAHIGTTTASTSALKIPQGVAFRAAAA 1564
QY 1615 AAHTTDTAGYNAMESVSREKPELTASTERVAKRMSVMVYSGLTPEEFMLVYKFAKHHI 1674
Db 1565 AGAD-----KAVVGIVSVKIKPELTSEERADRIDSMVYSGLTPEKVTMVQKFAKYRL 1617
QY 1675 TLNLTETEETHVVMKTDAEFVCERTILKYFLGIAGGKVVVSYFWYQTSIKERKMLNEHDF 1734
Db 1618 TLTDAITEETHVVIITDAEFVCERTILKYFLGIAGGKVIIVSYSWVVRISIQERRLLNVHEF 1677
QY 1735 EVRGDVYNGRNHOGPKRARESODRKIFRGLGICCYGPFNTNKPDDOLEBMVYQLCGASVYKE 1794
Db 1678 EVKGDVYNGRNHOGPKRRESRE-KLFGQLGVYCCPEFTNMPKDELRLMLQLCGASVYKE 1736
QY 1795 LSSFTLGTGVHPHVVVQPDADWEDNGFHAICQMCEAPVTVREWLVDLSVALYQCELDLYL 1854
Db 1737 LPSLTHDTGAHLVIVQPSAWTEDSNCPDGLQCLKARLVNMDVLDLSLYRCRDLDAYL 1796
QY 1855 IPQI 1858
Db 1797 VQNI 1800
```

RESULT 9
O54952

```
ID O54952 PRELIMINARY; PRT: 1817 AA.
AC O54952; p97951;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer
susceptibility protein).
```

```
GN BRCA1.
OC Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=99111388; PubMed=9892727;
RA Bennett L.M., Brownlee H.A., Hagavik S., Wiseman R.W.;
RT "Sequence analysis of the rat brca1 homolog and its promoter region.";
RL Mamm. Genome 10:19-25(1999).
RN [2]
RP SEQUENCE OF 8-222 FROM N.A.
RC STRAIN=WISTAR-KYOTO; TISSUE=SPLEEN;
RX MEDLINE=96358532; PubMed=8761410;
RA Chen K.S., Shepel L.A., Haag J.D., Heil G.M., Gould M.N.;
RT "Cloning, genetic mapping and expression studies of the rat Brca1
gene.";
RL Carcinogenesis 17:1561-1566(1996).
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION, MAY PLAY A ROLE
IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR EMBL; AF036760; AAC36493.1; -.
DR EMBL; S82504; -; NOT_ANNOTATED_CDS.
DR EMBL; S82502; -; NOT_ANNOTATED_CDS.
DR EMBL; U60523; AAB40387.1; -.
DR EMBL; S82500; AAB37501.1; -.
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR002378; Brst_cancer1.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00533; BRCT; 2.
DR Pfam; PF00097; zf-C3HC4; 1.
DR PRINTS; PR00493; BRSTCANCER1.
DR SMART; SM00292; BRCT; 2.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00172; BRCT; 2.
KW Zinc-finger; DNA-binding; Nuclear protein; Anti-oncogene.
FT ZN_FING 24 64 C3HC4-TYPE.
FT DOMAIN 497 503 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT CONFLICT 38 38 Q -> K (IN REF. 2).
FT CONFLICT 192 192 A -> M (IN REF. 2).
SQ SEQUENCE 1817 AA; 199876 MW; C0B4760F0E349A01 CRC64;
```

Query Match 50.5%; Score 4873.5; DB 11; Length 1817;
Best Local Similarity 56.2%; Pred. No. 2.9e-243;
Matches 1049; Conservative 259; Mismatches 488; Indels 69; Gaps 32;

```
QY 1 MDLSALRVEEVQNVINAMQKILECPICLELIKBPVSTKCDHIFCKMCLKLLNQKGPQ 60
Db 1 MDLSAVRIQEVQNVLIAMQKILECPICLELIKBPVSTQCDHIFCKMCLKLLNQKGPQ 60
QY 61 CPLCKNDITKRSLQESTRFSQVLEELLKIICAFQDITGLTGLYANSYFNAKKENNSPEHLK 120
Db 61 CPLCKNEITKRSLQSGARSQVLEELLKIIDAIFELDTGMOCANGSEFSKKNSSSELLNE 120
QY 121 EVSIIQSMGYRNRAKRLLOSEPENPSLQETSLSVQLSNLGTVTLTKQRIQPKTSVYI 180
Db 121 DASIIQSVGYRNVRKVLQIQESGATLKD-SLSVQLSNLGIIVSRMKNKRNQTPONKSVYI 179
QY 181 ELGSDSSEDTVNKATYCSVGDQELQITPOGTRDEISLDSAKKAACEFSFSDVTNTEHQ 240
Db 180 ALESDSSEERVNAPDGCVSVDQELFQIAPGAGDEGKLSAKKAACDFSE-GRNIEHHQ 238
QY 241 PSNNDLNTEKRAAERHPEKYQGSVSNLHVEPCGNTNTHASSLQHENSLLLTQKRMNVE 300
Db 239 CSDKDLNPTENHATERHPEKCPISVANVHVEFCGTDARASSLQRTGRTSLTLFTEDRLDAE 298
QY 301 KAEFCNKSQOPGLARSOHNRWAGSKETCNDRRPTSTPEKKYVDLNADPLCEKRNKQKLP 360
Db 299 KAEFCDRSKQSGAAVSQQRWADSKETCNGRPVPRTEGKADPNVDSLGRKQWNHPSKSLC 358
```

```
QY 361 SENPRDTEVPWITLNSSIQKYNFERSRDELIGSDSDHGESESNKAVADVLDVNEVD 420
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 359 PENSNGATTVPWITLNSSIQKYNFERSRDEGMTSDNASDRPASPANAASVLEVSNEVD 418
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 EYSGSSEKIDLLASDPHEALICKSERVSHKSVSYESNIEDIKFGKTYRKKASPLNLSHVTE 480
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 419 GCFSSKKIDLAVDPDNVADTSGRDESPVENIINDIKFGKTYRKGRSPLNHNVE- 477
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 LIICAFVTEPQIIQERPLTNKLRKRRTSGLHPDEFKKAIDL-AVQKTPEMINGCTWQT 539
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 478 -IIGFTTTEPQIIQEQPFTNKLKRR--STCLHPDEFKKAIDLTVYQRISENLNQTDQM 534
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 540 EONGQVANTINSNGENKTKTQNEKNPNPIESLEKESAPTKAEPISSTSSNMELELN 599
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 535 EPNDQAMSTISNGQENRATGNDLQRCRNAPHPESLKEPAFTAKAKSINSISDLEVELN 594
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 600 IHNSKAPKNRURRSSTRHIIHALELVVSRNLSPPNCTELQIDSCSSSEETKKKYNQMP 659
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 595 VHSKAPKNRURRK-STRCVLPLE-PISRNPSPPTCAELQIESCGSSEETKKNNSQTP 652
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 660 VRHSRNLQMECKEPATGAKSKNKNQETSKRHSDTTPPELKLITNAPGSFYKCSNTSELK 719
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 653 AGHIREPQIUEDTEPAADAKK-NEPNEHIRKRSASDAFPEEKLMMKAGLLTSCSPRPKP 711
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 720 EFVNPSPREEKEKLETVKVSNNABDPKDLMSGE-RVLOQERSVESSSISLVPQTDYG 778
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 712 GPVNPSP-PERKIGIEQLEMCOMPNNKELGDIVLGEPSGKPTPEPSESTSVSLVPTDID 770
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 779 TQBSISLLEVSTLGAKTENPKVCVSOCAFENPKGLIHGCKDNDRNDETFKYPILGHEVN 838
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 771 TONSVSILEANTVRYARTGSVCQMTQFVASENPKNELVHG-SNNAGSGSECFKHLRHELN 829
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 839 HSRETSIEMSESELDAQYLQNTFKVSKQSFALFSPNPGNAEEBCATFSAHSGSLKKQSPK 898
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 830 HNOET-IEMEDSELDTQYLQNTFQVSKQSFALFSLKRSPOKDCTLVGARSVPSPREPSPK 888
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 899 VTFPECKEENOGKKNESNKKPVOTNITAGPPVGVGOKPVDNNAKCSIKGGSRFCLSQF 958
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 889 VTSRGQK-ERQGESELSHVQAVTVVGLVPVQCEKPGAVTMCA--DVSRLCPSSH 945
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 959 RGNETGLITPNKHGLQNPYRPIPLPFKSVFKTKCKNLLLEENFEHSMSPEREMNGEN 1018
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 946 RSCENGLNTDKSGISQNSHFQSVPLRSSIKTDNRKLTLEGREKUT---ERGMGNET 1002
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1019 -IPSTVSTISRNIENYFKEASSNINEVGSSTNEVGSSINEIGSSDENTQOAEGLNRNG 1077
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1003 AVOSTIHTISLNN-RGDACLEASS-----GSVIEVHSTGENYOGOLDNRNG 1047
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1078 PKLNAMRLGLVQLPEYVYKQSLPGSNCKHPKIKQBYEEVQTVNTDFPYLISONLEQPM 1137
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1048 PKVNTVSLDSTQPGVSKQSAVSD-KYLEIKQES-----KAVSADFSPCLFSHILEKPM 1101
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1138 GSHASQVCSPTDDLLDGETIKETSPAENDIKESSAVFSKQVKGELSRSPSPFTHH 1197
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1102 RSDKTFQVCSEPTDDLLDVEIQENASGEGGITEKSAIFNGSVLRRSSRSPSPVTHAS 1161
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1198 LAGGYRRGAKKLESSEENLSSEDELPQFQHLFGKVNNIPSOSTRHSTVATECLSKNTE 1257
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1162 KRSULHRGSRKLEFSEESDSTDEDLPCFQHLL-SRVSSSTP-ELTRCSSVYTVRPEKAK 1219
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1258 ENLLSLKNSLNDSCNQVILAKASQPHLSSEETKCSASFSSQCSLEDLTANTQDPFL 1317
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1220 GTQAPRKSSIDCNNEVILGEASQEQVSEDAKCSGSMFSSQHSAAALSPANALSQDPDF 1279
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1318 IGSSQMRHQRSQGVGLSDKELVSDDEERTGLLEE-NNQEQSM--DSNLGEAASGES 1374
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1280 NPPSKORRHOAENEAFSLSKELISDHEDMAACLEASQDEEDSIIPDS-----VASGVES 1335
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1375 ETSVSDCSGLSSQSDILTTQORDTMOHNLKLOQEMAEVLEOHGSOQSPSNYPSTIS 1434
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1336 EANLSEDC-----SQSDILTTQORDTMOHNLKLOQEMAEVLEOHGSOQSPSGHPPLCPA 1391
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 1435 DSSALEDLRNPQSTSEKAVLTSSQSSSEYPISONPEGLSADKFEVS-ADSSSTSKNKEPGV 1493
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1392 DPCALEDLDPDEQNRSGTALTTSKNINENPVSONPKRACDDKASQPPDGLSPGDKESGM 1451
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1494 ERSSSKCPSLDDRWYMHSCSGSLQNRNYPQOEELIKVVYVEEQOLESPPHDLTETSYL 1553
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1452 RPSFPKSPLTSSRCSARGHSRSLQNRNSTSQEELLQPAKLE---KSCPHNLTRGSL 1507
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1554 PRODLGCTPYLESGLSLEFSDPESDRAPESARVGNIPSSTSALKVPOLKVAESAQS 1613
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1508 PRODLGCTPYLESGLSLEFSDPESDRAPESARVGNIPSSTSALKVPOLKVAESAQS 1564
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1614 PAAAHHTDTAGYNAMESVREKPELTASTERNRMSMVVSGLTPPEEMLVYKPAKHH 1673
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1565 PAAG-GADT---AVVEIVSKIKPEVTPSKERAERDISMVVSGLTPKPEVMIQKPAEKYR 1619
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1674 ITLNLITEETHVVMKTDABFVCERTLYKFLGIAGGKMWVSFYVWTSQIKERKMLNEHD 1733
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1620 LALTDVITEETHVVIKTDAEFVCEKTLKYLFLGIAGGKMWVSFYVWTSQIKERKMLNEHD 1679
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1734 FEVRGCVVNGRNHOGPKRARESODRKIFRGLICCYGPTNNMPTDOLEHMMVOLCGASVVK 1793
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1680 FEVRGCVVNGRNHOGPKRARESODRKIFRGLICCYGPTNNMPTDOLEHMMVOLCGASVVK 1738
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1794 ELSSTFLGTGVHPIVVVQPDADWNTDNGFHAIGOMCEAPVVTREWVLDVSVALYQCOBLDTY 1853
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1739 ELPLLTRDTGAHPVLVQPSAWNTEDNDCPDICQLCKGRVMDWVLDVSISVRCRDLAY 1798
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1854 LIPQI 1858
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1799 LVQNI 1803
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
O46490 PRELIMINARY; PRT: 1146 AA.
AC O46490;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Breast cancer type I susceptibility protein (Breast and ovarian cancer
DE susceptibility protein) (Fragment).
GN BRCA1.
OS Cynoccephalus variegatus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Dermoptera; Cynocephalidae; Cynoccephalus.
OX NCBI_TaxID=9457;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98122577; PubMed=9462745;
RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,
RA Fodor S.P.A., Brody L.C., Collins F.S.;
RT "Evolutionary sequence comparisons using high-density oligonucleotide
RT arrays.";
RL Nat. Genet. 18:155-158(1998).
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE
CC -!- IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
DR EMBL: AF019081; AAC39589.1; -.
DR InterPro: IPR002378; Brst_cancerI.
DR PRINTS: PR00493; BRSTCANCER1.
KW DNA-binding; Nuclear protein; Anti-oncogene.
FT NON_TER 1
FT DOMAIN 276 283 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 818 821 POLY-SER.
FT NON_TER 1146 1146
SQ SEQUENCE 1146 AA; 128330 MW; FD22388B602380AE CRC64;

Query Match 45.5%; Score 4386.5; DB 6; Length 1146;
Best Local Similarity 76.2%; Pred No. 2.2e-218;
Matches 874; Conservative 100; Mismatches 166; Indels 7; Gaps 6;

QY 225 ACEFSETDVTNTEHHQPSNNDLNTEKRAERHPKEKYOGSSVSNLHVPCGTNTHASSLQ 284
```

Db 1080 AVEDLATNTQDPFLMFDPSPNHNHRHQSENVVLSDKRELVSDDERETALEYNHEEE 1139

QY 1359 QSMDSNL 1365

Db 1140 QSVDSNL 1146

RESULT 11

O46489 PRELIMINARY; PRT; 1141 AA.

AC O46489;

DT 01-JUN-1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer susceptibility protein) (Fragment).

GN BRCA1.

OS Galago crassicaudatus (Thick-tailed galago) (Otolemur crassicaudatus).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Strepsirhini; Galagonidae; Otolemur.

OX NCBI_TaxID=9463;

RN [1]

SEQUENCE FROM N.A.

RP MEDLINE=98122577; PubMed=9462745;

RX Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,

RA Fodor S.P.A., Brody L.C., Collins F.S.;

RT "Evolutionary sequence comparisons using high-density oligonucleotide arrays.";

RL Nat. Genet. 18:155-158(1998).

CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

DR EMBL; AF019080; AAC39588.1; -.

DR Interpro; IPR002378; Brst_cancer1.

DR PRINTS; PRO0493; BRSTCANCER1.

KW DNA-binding; Nuclear protein; Anti-oncogene.

FT NON_TER 1

FT DOMAIN 277 284 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

FT DOMAIN 199 202 POLY-SER.

FT DOMAIN 816 819 POLY-SER.

FT NON_TER 1141 1141

SQ SEQUENCE 1141 AA; 127578 MW; 2653CA894834946F CRC64;

Query Match 41.9%; Score 4040; DB 6; Length 1141;

Best Local Similarity 72.1%; Pred. No. 1.7e-200;

Matches 831; Conservative 105; Mismatches 192; Indels 24; Gaps 15;

QY 226 CBFSETDVTNTEHHQPSNNDLNTTEKRAAEHRPEKYQGSVSNLHVPCGTNTHASSLQH 285

Db 2 CMVSEKDIITSIEHQSSNKDLNTIEKRATEMHPEKYQGSVSNLHVPCGKNTNTHASSLQH 61

QY 286 ENSSLLLTQDRNVEKAEFCNKSQKPLARSQHNWAGSKETCNDRRTSTSEKKVLDLND 345

Db 62 ESSSLLLTQDRNVEKAEFCNKSQKPLARSQHNWAGSKETCNDRHTCSPQKVDLNTA 121

QY 346 PLCEKREKWKQKLPCSENPRDTEDYPTLNSIQKVNWFSDLLGSDSDHGESES 405

Db 122 PPGYRKEQKREKLLCSKNPRSDYPTLNSIQKVNWFSDSDHGESES 181

QY 406 NAKVADVLVLEVDYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTY 465

Db 182 HAEVAGALEVPEVDGYSSSEKIDLLASDPHYPIICKSERVHSKPIKSKVEDKIFGKTY 241

QY 466 RKKASLPNLSHVTEENLIIGAFVTEPQIIQERPLTNKLKRKRPTSGLHPEDFIKKADLA - 524

Db 242 RKKASLPNLSHVTEENLIIRAAATEPQITQECSLTNKLKRKRRTTSGLCPEDFIKKADLAV 301

QY 525 VOKTPEMINQGTNQTQONQOVNITNSHENTKGDSONEKNPNPIESLEKESAFKTKA 584

Db 302 VOKTPEKRIQGTNOVDONSHVNITNSGYENETKGDYQVONEKNANSTESLEKESLGTKA 361

QY 585 EPTSSSISNMELELNTHNSKAPKKNRLKRKSTRHHALELVVSRNLSPPNCTELQIDSC 644

Db 1 ACDFSE-DITNIEHQSCDKDLNTIEKHATERHPPEKYQSSVSNLHVPCGTNTHASSLQ 59

QY 285 HENSLLLTQDRNVEKAEFCNKSQKPLARSQHNWAGSKETCNDRRTSTSEKKVLDLNA 344

Db 60 HENSLLLTQDRNVEKAEFCNKSQKPLARSQHNWAGSKETCNDRRTSTSEKKVLDLNA 119

QY 345 DPLCEKREKWKQKLPCSENPRDTEDYPTLNSIQKVNWFSDLLGSDSDHGESE 404

Db 120 DSQGRKERKWKQKPPYSPESPRDTQDPVPTLNSIQKVNWFSDLLGSDSDHGESE 179

QY 405 SNAKVADVLVLEVDYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKT 464

Db 180 SNAKVAGALEVPEVDGYSSSEKIDLLASDPHDALICKSERVHSKPIKSKVEDKIFGKT 239

QY 465 YRKASLPNLSHVTEENLIIGAFVTEPQIIQERPLTNKLKRKRPTSGLHPEDFIKKADLA 524

Db 240 YRKASLPNLSHVTEENLIIGAFVTEPQIIQERPLTNKLKRKRPTSGLHPEDFIKKADLA 299

QY 525 -VOKTPEMINQGTNQTQONQOVNITNSHENTKGDSONEKNPNPIESLEKESAFKTK 583

Db 300 VOKTPEKINOIQDQONQOVNITNSHENTKGDYQVONEKNANSTESLEKESAFRTK 359

QY 584 AEPISISSNMELELNTHNSKAPKKNRLKRKSTRHHALELVVSRNLSPPNCTELQIDS 643

Db 360 AEPISISSNMELELNTHNSKAPKKNRLKRKSTRHHALELVVSRNLSPPNCTELQIDS 419

QY 644 CSSEBEIKKKYQKQPVNRSNRLQMEGEPATGAKSKNPNPQTSKRHSDSTFPELKL 703

Db 420 CSSEBEIKKKYQKQPVNRSNRLQMEGEPATGAKSKNPNPQTSKRHSDSTFPELKL 479

QY 704 NPGSFTKCSNTSELKEFVNPSLPREEKEKLETVKVSNAEDPKDLMLSGERVLOTERS 763

Db 480 NVPVVFANCSNKKLQEFIDPSLQREIEENLETIHVSNSAKDPKDLVLSGEGLOTERS 539

QY 764 VESSISLVPGTDYGTQESISLLEVLTKAKTEPNKVCYQCAAFENPKGLIHGCSKDN 823

Db 540 VESTSISLVPGTDYGTQESISLLEVLTKAKTEPNKVCYQCAAFENPKGLIHGCSKDN 599

QY 824 NDEGKPYPLGHVNHUSRETSIEMSELDQAQYQNTFKVSRQSFALSPNCGNAEECA 883

Db 600 NDEGKPYPLGHVNHUSRETSIEMSELDQAQYQNTFKVSRQSFALSPNCGNAEECA 659

QY 884 TFSAGSGLKKQSPKVTFCQEQEENQKNEISNKPQVNTITAGPVPVVGQK-DKPVDNA 942

Db 660 TVYAHSLRLKQSPKVTFCQEQEENQKNEISNKPQVNTITAGPVPVVGQK-DKPVDNA 719

QY 943 KCSIKGGRSFCJSSQFRGNETGLITPNKHGLLQNPYRIPPLFPKIFSVKTKCKNLEEN 1002

Db 720 KFSIKGGRSFCJSSQFRGNETGLITPNKHGLLQNPYRIPPLFPKIFSVKTKCKNLEEN 779

QY 1003 FEHSHSPEREMGNEN-IPSTVSTISRNIRNENVFKEASSNINEVGSSTNEVGSSINEI 1061

Db 780 FEHSHSPEREMGNEN-IPSTVSTISRNIRNENVFKEASSNINEVGSSTNEVGSSINEI 839

QY 1062 GSDENIQALGRNRPKLNAMLRVLQPEVYKQSLPCSNCKHPKIKKOEVEEVQTVN 1121

Db 840 GPGENIQALGRNRPKLNAMLRVLQPEVYKQSLPCSNCKHPKIKKOEVEEVQTVN 899

QY 1122 TDFSPYLIDNLEQPMGSSHASQVCSSETPDDLLDGDGEIKEDTSFAENDIKESSAVFSKV 1181

Db 900 TNLCLISHLNLEQPMGSSHASQVCSSETPDDLLDGDGEIKEDTSFAENDIKESSAVFSKV 959

QY 1182 QKGLSRSPSPFTHLAQYRGAKKLESSEENLSSDEELPCFOHLLFGKVNNTIPQS 1241

Db 960 QKGLSRSPSPFTHLAQYRGAKKLESSEENLSSDEELPCFOHLLFGKVNNTIPQS 1019

QY 1242 TRHSTVATECLSKNTEENLLSKNSLNDSCNOVILAKASOEHLSEETKCSASLFSQCS 1301

Db 1020 TRHSTVATECLSKNTEENLLSKNSLNDSCNOVILAKASOEHLSEETKCSASLFSQCS 1079

QY 1302 ELBDLTANTNTQDPFLI--GSSQKMRHQSESQGVGLSKRELVSDDERETALEYNNQEE 1358

```
Db 362 EPISSISNMKLELNTHNSKASKKKRLRKKSSRHIRALELVVNNKPPNHNTHLIQIDSC 421
QY 645 SSSEIEIKKKYINOMPVRHNRNLQMECKBPATGAKKSNKPNBOTSRRHSDSTFPELKLNTN 704
Db 422 SSSEIEIKKSSDOIPIVRHSRKPCLMEDREPATGAKKSNKPNBQISKRHVSDTFPEVALTN 481
QY 705 APOGFTKCSNTSELKEFVNPSLPREKEKL-ETVKVSNNAAEDPKDMLSGERVLLQ---T 760
Db 482 ISSFFTNCSGSNRLKEFVNPSLQRKTEENLETTIQVSNSTKGP---VLSGERVLIQISE 538
QY 761 ERSEVSSSTSLVPGTDYGTQBSISLLEVSTLCKAKTEPNKCVSQCAAFENPKGLIHGCSK 820
Db 539 ERSIKSTSLVDPDITDYGTDNSLKVLRKVTAPNKHASQGTATENPKELIHGCSK 598
QY 821 DNENDTEGFKYPLGHEVNSHRSSTSEMESELDQYLONTFKVSKRQSFALFSPGNAAE 880
Db 599 DTGNDTEGKYDPLRHEINHIQISMEDESELDQYLONTFKFSKQSFALFSLNIG---K 655
QY 881 ECATFSAH---SGSLKKQSPKVTFECEQKEENOGKNESNIKPVQTVNITAGFPVVGQDKP 938
Db 656 ECATVCAQSLSASLRKGS-KVILECEQ-IENPGMKPEKIKHTQGNININTGFSVVCQDKK 713
QY 939 VDN-AKCSITKSGSRCLSSQFRGNETGLTPNKHGLLQNPRIPLPIKSVFTKCKKN 997
Db 714 TDDYAKYSIKEASRFLCSNQFRDNETESITVKNLGILOQLYHIPPLSPIRLEFDKTCNTN 773
QY 998 LLEENFEHSMSPEREMGNEN-IPSTVSTISRNNIRENVEKEASSNINEVSGSTNEVGS 1056
Db 774 LLEEREHSHVLPEKAVGNENTVPMTMTIIONN-RESAYKEASSSSINEVSSSTNEVGS 832
QY 1057 SINEIGSSDENTQAELOELGRNRGPKLNAVLRLGLVQPEVYKQSLGNSCKHPKIKQYEYEV 1116
Db 833 SYNEVGPSSENTQAELOELGRGPKLNAVLRLGLMQPEVYKQNLPSNCEHPKIKQEEYEV 892
QY 1117 VQTVNTDFSPYLLISNDLEQPMGSSHASQVCSSETPDDLLDDGEIKEDTFAENDIKESAV 1176
Db 893 VQVPNPDFSCSLISNDLEQPTRESSHASQVCSSETPDDLLVDDDELKENTSFABNNIKERSA 952
QY 1177 FKSVOGGLSRSPSPFHTHLAQCYRRGAKKLESSEENLSSEDEELPCFOHLLFGKVN 1236
Db 953 FSKNVMREISRSPSLAHLHTQAHQREVRLKLESSEENMSSEDEELPCFOHLLFGKVTN 1012
QY 1237 IPSQSTRHSTVATECLSKNTENLNLNSLNSLNDSCNQVILAKASQEHLSBETKCSASLF 1296
Db 1013 TSSQSTROSTIATECLSRKTEETLVSQNTLNDSCNQVTLVKASQENHLSSEAKCSVSFL 1072
QY 1297 SSQCSLEDLTANTTQDPLIGS--SKOMRHQSESQVGLSKDELVSDDERGTGLEE- 1353
Db 1073 SSQCS--EDLTANTTQDPLMSDPLSKOMRDQSENQEV-LSDNELVSNDEEREPSLEED 1129
QY 1354 NNOEQSDMSNL 1365
Db 1130 NHQEQSVSDSL 1141
RESULT 12
Q924E0 PRELIMINARY; PRU; 1063 AA.
AC Q924E0.
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE BRCA1 (Fragment).
GN BRCA1.
OS Pedetes capensis (Springhaas).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Pedetidae; Pedetes.
OX NCBI_TaxID=10023;
RN [1]
RP MEDLINE=21219190; PubMed=11319262;
RX Adkins R.M., Gelke E.L., Rowe D., Honeycutt R.L.;
RA "Molecular phylogeny and divergence time estimates for major rodent
RT
```

```
RT groups: evidence from multiple genes.;
RL Mol. Biol. Evol. 18:777-791(2001).
DR EMBL: AF332047; AA71627.1; -.
FT NON_TER 1 1063
FT SEQUENCE 1063 AA; 118726 MW; 9B9CE758FE40745 CRC64;
SQ
Query Match
Best Local Similarity 69.5%; Score 3693; DB 11; Length 1063;
Matches 752; Conservative 103; Mismatches 203; Indels 24; Gaps 7;
QY 262 QGSVSNLHVPCGTNTHASSLQSHENSLLLTKDRMNVKEAFPCNKSQOPGLARSQHNRW 321
Db 1 QGISVSNLCVPCGTNTHASSLQSHENSLLLTKDRMNVKEADFCNKSQOPGLARSQHNRW 60
QY 322 AGSKTCNDRTPTSTKVKYLDNADPLCERKEWKKOKLPCSENPRDTEYDWPITLNSSTQK 381
Db 61 AESKETCNDRTPTSTKVKYLDNADPLFYGRKERKKOKPPCSENPRDQVSWYTLNSSTQK 120
QY 382 VNEWFSRDELGLGSDSHDGESESNKAVDVLVNEVDEYSGSSEKIDLLASDPHEALI 441
Db 121 VNEWFSRDEMLTSDDLHDGVSESNAEVLGALEVPDEIKHGSGSEKIDLLASDPHNALT 180
QY 442 CKSERVSHKSVESNIEDKIFGKTYRKASLNSHVTENLIIGAFVTEPQIIEERPLTNK 501
Db 181 CKSERVCAKPVESNIEDKVFGRKISLNSLSPITENLVGAFATKPIQRQEHPLTNK 240
QY 502 LKRRRTSLGHPDEFTKKADLAV-OKTPEMINOGTNOEQGVMMNTNSGHEKNTKGD 560
Db 241 LKRRKATSLRPEDFTKKVDLTVIQKTEKINOCTAQMGQGVQVNNNTNLENETKND 300
QY 561 SIQNEKNPNPTESLEKESAFKTKAEPITSSISNMELENIHNSKAPKNRRLRRKSSRHI 620
Db 301 YVLKEKNANPLESLEKESRTHRTKAEPITSSISNMELENIHNSKAPKNRRLRRKSSRHI 360
QY 621 HALELVYSRLNSPPNCTELQIDSCSSSEIEIKKKYINOMPVRHNRNLQMECKBPATGAKK 680
Db 361 HALELVVNNKPPNPNQTELQIDSCSSSEIEIKKKYINOMPVRHNRNLQMECKBPATGAKK 420
QY 681 SNKNEQTSKRHSDSTPELKLITNAPGFTKCSNTSELKEFVNPSLPREEKEKLETYKV 740
Db 421 NNTVEKISEKHATDAPPEVKVTNIPGLNNCNSNKPPEEFVSPGFQREIEENLETQV 480
QY 741 SNAEDPKDMLSGERVLLQTERSVESISSISLVPGTDYGTQBSISLLEVSTLCKAKTEPNK 800
Db 481 SNSTKDKNLVLSAEKGLPPERSTESTSVSLVPTDYDQDSISLLETNLTLRKAKTMSNQ 540
QY 801 CVSOCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNSHRSSTSEMESELDQYLONT 860
Db 541 CMTQYVAIEHPKELSHSCSKDTRNDTKGFKDALRREVNHIQETNTEMESELDQYLONT 600
QY 861 FKYSKQSFALFSPGNAAEBCATFSAHSGSLKKQSPKVTFECEQKEENOGKNESNIKPV 920
Db 601 FOAKSKQSFALFSPNRNPEKEC-----AHSMSLRKQSTKVTLEYEQKEENOGKESKIKPA 656
QY 921 QTVNITAGFPVVGOKDKPVDNAKCSIKGGSRCFLSSQFRGNETGLITPNKHGLLQNPYRI 980
Db 657 QEYNTTVGFPVWCOEDKPCADANHSITGVSRCLPCLSQLTDNETELSTVNNKHGVLPNPLYI 716
QY 981 PPLFPKISFYKTKCKKNLLEENFEHSMSPEREMGNEN-IPSTVSTISRNNIRENVEFKEA 1039
Db 717 PSISPVKSFITTKYKSPSEBEKFKYSASHENAMGNESIIQSIIVYTIQNNIRENAFKEA 776
QY 1040 SSSNINEVSGSTNEVSGSSINEIGSSDENIQAEGLCRNGPKLNAVLGLVQPEVYKQSLP 1099
Db 777 SS-----GSINEVSGSMMQVLEGRNGPKLNSVQRGLHQIQRKQSLP 822
QY 1100 GSNCKHPKIKK-QEYEEVWQVTNDFSPYLLISNDLEQPMGSSHASQVCSSETPDDLLDDGE 1158
Db 823 VSS-KHPKEMKKQENEGVQAVDAFSPCLTSNSLEQPMRSGCVQICSETPDDLLDDGE 881
QY 1159 IKEDTSAENDIKESSAVFSKQVKGELSRSPFTHTHLAQCYRRGAKKLESSEENLS 1218
Db 1159 IKEDTSAENDIKESSAVFSKQVKGELSRSPFTHTHLAQCYRRGAKKLESSEENLS 1218
```

```
Db 882 VKENTSTEGIGKETS AVSKFSIQRGESSRSPPLTHISLQAQCHRRARKLESSESISS 941
QY 1219 EDEELPCFQHLFGKVNIPQSQRHSTVATECIGLSKNTNENLLSLKNSLDCSNQVILAK 1278
Db 942 EDEELPCFQHLFGKVTNTSQSRGNNAVTEYLSKTKENLVSLKTDLDNCNDLILAK 1001
QY 1279 ASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLI--GSSKMRHQSSQGVGLS 1336
Db 1002 TSOEPLHSEEARYSGLFSQCSVLEDLPLANTNTQSSFLMFGPPSKQTRHQSENQEVLS 1061
QY 1337 DK 1338
Db 1062 DK 1063

RESULT 13
Q8WMT3 PRELIMINARY; PRT; 947 AA.
AC Q8WMT3;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE BRCA1 (Fragment).
GN BRCA1.
OS Tapirus pinchaque (Mountain tapir).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Tapiridae; Tapirus.
OX NCBI_TaxID=30582;
RN [1]
RP SEQUENCE FROM N.A.
RA Murphy W.J., Eizirik E., O'Brien S.J., Madsen O., Scally M., Douady C.,
RA Teeling E., Ryder O., Stanhope M.J., deJong W.W., Springer M.S.;
RT "Resolution of the early placental mammal radiation using Bayesian
RT Phylogenetics.";
RL Science 0-0-0(2001).
DR EMBL; AY057830; AAL30077.1; -.
FT NON_TER 1
FT NON_TER 947
SQ SEQUENCE 947 AA; 104734 MW; 21D8DE08EBDE0D43 CRC64;

Query Match 36.9%; Score 3557.5; DB 6; Length 947;
Best Local Similarity 75.1%; Pred. No. 1.4e-175;
Matches 713; Conservative 79; Mismatches 152; Indels 5; Gaps 5;

QY 274 CGTNTHASSLOHENSLLLTKDRMNYEKAFCNKSOPGLARSOHNRWAGSKETCNDRT 333
Db 1 CGTNTHASSLOHENSLLLTKDRMNYEKAFCNKSOPGLARSOHNRWAGSKETCNDRT 60
QY 334 PSTEKVKVDLADPLCERKEWKNOKLPCSENPRDTEVPWITLNSSIQKVNWFSSRDELL 393
Db 61 PNSEKKVLVNLADPLHGRKELNKKQPCSNSSRSDQDVPWITLNSSIQKVNWFSSRDEML 120
QY 394 GSDSDHGESESNKAVADVLDVLENDVEXSGSEKIDLLASDPHEALICKSERVHSKVE 453
Db 121 SDDSDRGGPTEHTEVAGAVEVNEVDVYSGSEKIDLMASDPGALICESERVHSKPVA 180
QY 454 SNIEDKIFGTYRKASLPNLSHVTENLIICAFVTEPQIITQERPLTNKLRKRRTSGHL 513
Db 181 NNIEDKIFGTYRKANLPNLSHITELIIEASAIEPQIITQERPLTNKLRKRRTSGHL 240
QY 514 PEDFIKKADLA-VQKPEMTNOGTNOTEQNGQVNMNTNSGHENKTKGDSIQNEKNPNPIE 572
Db 241 PEDFIKKVDPAIVQKPEKVEIEQTEQNDHVMNTIANNGHENETKGCVOKEKNANPTE 300
QY 573 SLEKSAFKTKAEPISSSISNMELELNHNKAPKNLRKRKSTRHIALELVVSRNLS 632
Db 301 SLEKSAFRTAEAPISSSISNLELNHNFRSSKAPKNLRKRKSTRHIALELVVSRNPS 360
QY 633 PPNCTELOIDCCSSSEIEKKKKYNOMPVRHSRLQLMGEGKEPATGAKKSNKPNQTSKRH 692
Db 361 PPNHTELOIDCCSSSEIEKKKNSQGPGRHRRKLQLMEDKAPATGAKKSNKPNQINKRL 420
QY 693 DSDTFPELKLTNAPGSFTKCSNTSELKFEVNPSPREEKEKLETYKVSNNAEPPDKDML 752
```

```
Db 421 ASHAFPELKLNTNPGFANCSSNKLQEFVNPSLQREIEQNLAGAQVFNSSKDPKDLIL 480
QY 753 SGERVLQTERSVESSISLVPGDYGTQESISLVEYSTLGKAKTEPNKCVSOCAAFENPK 812
Db 481 SSGRGLOAKESVESTISLVDPDYGTQDISISLEADPLGKAKTAPNOGASCAATENSK 540
QY 813 GLIHGCSKNRNDTEGKYPLGHEVNHRSRETSLEMESELDQAQYLQNTFKVSKRQSFALF 872
Db 541 ELMHGCSKDRNDEGLKDP LXRNVHTQEPSTIEMESELDTOYLHNTFKVSKRQSFALF 600
QY 873 SNPGNAEECATSFSAHSGSLKKQSPKVTTECQKEBNQGNKNESNIKPQVTNITAGFPVY 932
Db 601 SNSGN-PEECATVCAHSRPLRKESPKVTLECGRKEBNQGNKESKIKHVOSVHTPAGEFPV 659
QY 933 GOKDKPVDNAKCSIKGSRFCLSSQPRGNETGLITPNKHCILLQNPYRIPPLPKSFVK 992
Db 660 CQDKPDGYVKSLKEVSRQCQSQPRGNETELITNKHGLSONLYIYIFSPSIRSVKT 719
QY 993 KCKKNLLEENFEHSMSPEREMGNEN-IPSTVSTISRNINRENVFKEASSSNIINEVGSST 1051
Db 720 TCKKNVSEKLEHESVSPERAXGNESIQTSTMTISQNNIRESTFEVGSSTINEVGSST 779
QY 1052 NEVGSSINEIGSSDENIQAEELGRNRPKNAMLRVLQPEVYKQSLPGSNCKHPKIKKO 1111
Db 780 NEVGSSINEVSSGENIQAEELGRNRPKNAILRLGLMQPEVYKQSHPISNCKHPKIKRO 839
QY 1112 -EYEEVVQVTNTDPSLYLISDNLEQPMGSSHASQVSETPDDLLDDGEIKEDTSPAENDI 1170
Db 840 GEMEGVVQAVNTDPSCLISDNLEQPMGSS-CAQVCSETPDDLLNDKDKIEENINFAESGI 898
QY 1171 KESSAVFSKSVQKGLSRSPSPETHLAQGYRRGAKKLESSEENLSSE 1219
Db 899 KERSAVFSKSVQKGFGRSPSPAHTCLAQGHORGARKLESSEENMSSE 947

RESULT 14
Q9BDV3 PRELIMINARY; PRT; 947 AA.
AC Q9BDV3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE BRCA1 (Fragment).
OS Dicerus bicornis (Black rhinoceros).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Rhinocerotidae; Dicerus.
OX NCBI_TaxID=9805;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21082081; PubMed=11214318;
RA Madsen O., Scally M., Douady C., Kao D., DeBry R., Adkins R.,
RA Amrine H., Stanhope M., de Jong W., Springer M.;
RT "Parallel adaptive radiations in two major clades of placental
RT mammals.";
RL Nature 409:610-614(2001).
DR EMBL; AF284011; AAK15598.1; -.
DR InterPro: IPR002378; Brst_cancer1.
DR PRINTS; PR00493; BRSTCANCER1.
FT NON_TER 1
FT NON_TER 947
SQ SEQUENCE 947 AA; 104753 MW; BC8158CB0F9A404B CRC64;

Query Match 36.8%; Score 3555.5; DB 6; Length 947;
Best Local Similarity 75.2%; Pred. No. 1.4e-175;
Matches 714; Conservative 86; Mismatches 143; Indels 7; Gaps 5;

QY 274 CGTNTHASSLOHENSLLLTKDRMNYEKAFCNKSOPGLARSOHNRWAGSKETCNDRT 333
Db 1 CGTNTHASSLOHENSLLLTKDRMNYEKAFCNKSOPGLARSOHNRWAGSKETCNDRT 60
QY 334 PSTEKVKVDLADPLCERKEWKNOKLPCSENPRDTEVPWITLNSSIQKVNWFSSRDELL 393
Db 61 PNSEKKVLVNLADPLHGRKELNKKQPCSNSSRSDQDVPWITLNSSIQKVNWFSSRDEML 120
QY 394 GSDSDHGESESNKAVADVLDVLENDVEXSGSEKIDLLASDPHEALICKSERVHSKVE 453
Db 121 SDDSDRGGPTEHTEVAGAVEVNEVDVYSGSEKIDLMASDPGALICESERVHSKPVA 180
QY 454 SNIEDKIFGTYRKASLPNLSHVTENLIICAFVTEPQIITQERPLTNKLRKRRTSGHL 513
Db 181 NNIEDKIFGTYRKANLPNLSHITELIIEASAIEPQIITQERPLTNKLRKRRTSGHL 240
QY 514 PEDFIKKADLA-VQKPEMTNOGTNOTEQNGQVNMNTNSGHENKTKGDSIQNEKNPNPIE 572
Db 241 PEDFIKKVDPAIVQKPEKVEIEQTEQNDHVMNTIANNGHENETKGCVOKEKNANPTE 300
QY 573 SLEKSAFKTKAEPISSSISNMELELNHNKAPKNLRKRKSTRHIALELVVSRNLS 632
Db 301 SLEKSAFRTAEAPISSSISNLELNHNFRSSKAPKNLRKRKSTRHIALELVVSRNPS 360
QY 633 PPNCTELOIDCCSSSEIEKKKKYNOMPVRHSRLQLMGEGKEPATGAKKSNKPNQTSKRH 692
Db 361 PPNHTELOIDCCSSSEIEKKKNSQGPGRHRRKLQLMEDKAPATGAKKSNKPNQINKRL 420
QY 693 DSDTFPELKLTNAPGSFTKCSNTSELKFEVNPSPREEKEKLETYKVSNNAEPPDKDML 752
```

```
Db 61 PNSEKLVNADPLVGRKELNKKQKPPCSDSPRDSQDIPWITRNSIQKVNWFSSDEIL 120
QY 394 GSDSDHGESESNKAVADVLVDNEVDEYSGSSEKIDLLASDPHEALICKSRVHSKSV 453
Db 121 TSDDSHDGGPESNTEVAGAVEQNEVDYSGSSEKIGLWASDPGALICESERVHSKPVE 180
QY 454 SNTIEDKIFGKTYRKASLNPNSLVHTENLIIGAFVTEPQIIQERPLTNLKKRRRTSGIH 513
Db 181 NNTIEDKIFGKTYRKASLNLNLSHITEDLIIGASATESQITQERPLTNLKKRRRTSGIH 240
QY 514 PEDFIKADLA-VOKTPEMINOGTNOEQOVNITNSCHENKTKGDSIQNEKNPNPTE 572
Db 241 PEDFIKADLA-VOKTPEMINOGTNOEQOVNITNSCHENKTKGDSIQNEKNPNPTE 300
QY 573 SLEKESAFKAPPISSISNMELENIHNSKAPKKNRLLRKSSTRHIALELVYSRNL 632
Db 301 SLEKESAFKAPPISSISNMELENIHNSKAPKKNRLLRKSSTRHIALELVYSRNL 360
QY 633 PPCTELQIDSCSSSEIEIKKKYNQMPVHRHNSLQIMEGKEPATGAKSKNPNQTSKRH 692
Db 361 PPNTHELQVDSCSSSEIEIKKKYNQMPVHRHNSLQIMEGKEPATGAKSKNPNQTSKRH 420
QY 693 DSDTFPELKLITNAPGFTKCSNTSELKEFVNPSLPREKEEKLTVKVSNNAEKPKDLM 752
Db 421 ASDAFPELKLITNAPGFTKCSNTSELKEFVNPSLPREKEEKLTVKVSNNAEKPKDLM 480
QY 753 SGERVLOTERSVESSISLVPGDTGTOESISLLEVSITLGRKAKTEPNKCVSOCAAFENPK 812
Db 481 SGERVLOTERSVESSISLVPGDTGTOESISLLEVSITLGRKAKTEPNKCVSOCAAFENPK 540
QY 813 GLIHGSKDNNDTEGFKYPLGHEVNSHRETSIEMEESLDAQYLQNTFKVSKRSFALF 872
Db 541 ELTHDCSKDTRNDTEGFKYPLGHEVNSHRETSIEMEESLDAQYLQNTFKVSKRSFALF 600
QY 873 SNGPNAEECATFSAHSGSLAKQSPKVTPECEOEKNOGKNESNIKPVQTNITAGFPV 932
Db 601 SNGPNAEECATFSAHSGSLAKQSPKVTPECEOEKNOGKNESNIKPVQTNITAGFPV 660
QY 933 GKQD-KPVDNAKCSITKGSFRCLSSQFRNETGLITPNKHGLLQNPYRIPPLFIKSPVK 991
Db 661 GKQD-KPVDNAKCSITKGSFRCLSSQFRNETGLITPNKHGLLQNPYRIPPLFIKSPVK 720
QY 992 TKCKKNLLEENFEHSMSPERMGNEI-IPSTVSTISRNIRENVFKKASSNINEVGSS 1050
Db 721 TCKKNVSEKLEHVSPEALGNKSVTQSTVSTISQNNIRENVFKKASSNINEVGSS 780
QY 1051 TNEVGSSINEIGSSDENIQAELGRNRPKLNAMLRLGLVQPEVYKQSLPGSNCKHPEIK 1110
Db 781 TNEVGSSINEIGSSDENIQAELGRNRPKLNAMLRLGLVQPEVYKQSLPGSNCKHPEIK 840
QY 1111 Q-EYEWVOTVNTDFSPYILISNLEQPMGSSHASOVGSETPDDLDDGEIKEDTSFAEND 1169
Db 841 RGENEGVQPVNADSPCPSISNLEQPMGSSHASOVGSETPDDLDDGEIKEDTSFAEND 900
QY 1170 IKSSAVFSKSVQKGLSRLSPFTHTHLAQYRRGAKKLESSEENLSSE 1219
Db 901 IKERSAIFSQVKGFEFRSPSPARTCLAQ--RGARKLESSEENMSSE 947

RESULT 15
Q9BDV4
ID Q9BDV4 PRELIMINARY; PRT: 942 AA.
AC Q9BDV4;
DT 01-JUN-2001 (TreeBLrel. 17, Created)
DT 01-JUN-2001 (TreeBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TreeBLrel. 18, Last annotation update)
DE BRCA1 (fragment).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_Taxid=9796;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=21082081; PubMed=11214318;
RA Madsen O., Scally M., Douady C., Kao D., DeBry R., Adkins R.,
RA Amrine H., Stanhope M., de Jong W., Springer M.;
RT "Parallel adaptive radiations in two major clades of placental
mammals.";
RL Nature 409:610-614(2001).
DR EMBL: AF284010; AAK15597.1; -.
DR InterPro: IPR002378; Brst_CancerI.
DR PRINTS: PR00493; BRSTCANCER1.
FT NON_TER 942 942
FT NON_TER 1 1
SQ SEQUENCE 942 AA; 104317 MW; F5C2EEE1C7313C1A CRC64;

Query Match 36.1%; Score 3482; DB 6; Length 942;
Best Local Similarity 73.6%; Pred. No. 9e-172;
Matches 698; Conservative 90; Mismatches 151; Indels 10; Gaps 4;

QY 274 CGTNTHASSLQHENSSLLLTKDRMNEKAEFCNKSQKPLARSOHNRWAGSKETCNDRT 333
Db 1 CGTNTHASSLQHENSSLLLTKDRMNEKAEFCNKSQKPLARSOHNRWAGSKETCNDRT 60
QY 334 PSTKKVVDLNADPLCERKEWNKOKLPCSENPRDTEDEVPWITLNSIQKVNWFSSDELL 393
Db 61 PNSEKLVNADPLVGRLEENKQKPPRSDSPRDSQDIPWITLNSIQKVNWFSSDELL 120
QY 394 GSDSDHGESESNKAVADVLVDNEVDEYSGSSEKIDLLASDPHEALICKSRVHSKSV 453
Db 121 TSDDSHDGGPESNTEVAGAVEQNEVDYSGSSEKIGLWASDPGALICESERVHSKPVE 180
QY 454 SNTIEDKIFGKTYRKASLNPNSLVHTENLIIGAFVTEPQIIQERPLTNLKKRRRTSGIH 513
Db 181 NNTIEDKIFGKTYRKASLNLNLSHITEDLIIGASATESQITQERPLTNLKKRRRTSGIH 240
QY 514 PEDFIKADLA-VOKTPEMINOGTNOEQOVNITNSCHENKTKGDSIQNEKNPNPTE 572
Db 241 PEDFIKADLA-VOKTPEMINOGTNOEQOVNITNSCHENKTKGDSIQNEKNPNPTE 300
QY 573 SLEKESAFKAPPISSISNMELENIHNSKAPKKNRLLRKSSTRHIALELVYSRNL 632
Db 301 SLEKESAFKAPPISSISNMELENIHNSKAPKKNRLLRKSSTRHIALELVYSRNL 360
QY 633 PPCTELQIDSCSSSEIEIKKKYNQMPVHRHNSLQIMEGKEPATGAKSKNPNQTSKRH 692
Db 361 PPNTHELQVDSCSSSEIEIKKKYNQMPVHRHNSLQIMEGKEPATGAKSKNPNQTSKRH 420
QY 693 DSDTFPELKLITNAPGFTKCSNTSELKEFVNPSLPREKEEKLTVKVSNNAEKPKDLM 752
Db 421 ASDAFPELKLITNAPGFTKCSNTSELKEFVNPSLPREKEEKLTVKVSNNAEKPKDLM 480
QY 753 SGERVLOTERSVESSISLVPGDTGTOESISLLEVSITLGRKAKTEPNKCVSOCAAFENPK 812
Db 481 SGERVLOTERSVESSISLVPGDTGTOESISLLEVSITLGRKAKTEPNKCVSOCAAFENPK 540
QY 813 GLIHGSKDNNDTEGFKYPLGHEVNSHRETSIEMEESLDAQYLQNTFKVSKRSFALF 872
Db 541 ELTHDCSKDTRNDTEGFKYPLGHEVNSHRETSIEMEESLDAQYLQNTFKVSKRSFALF 600
QY 873 SNGPNAEECATFSAHSGSLAKQSPKVTPECEOEKNOGKNESNIKPVQTNITAGFPV 932
Db 601 SNGPNAEECATFSAHSGSLAKQSPKVTPECEOEKNOGKNESNIKPVQTNITAGFPV 660
QY 933 GKQD-KPVDNAKCSITKGSFRCLSSQFRNETGLITPNKHGLLQNPYRIPPLFIKSPVK 991
Db 661 GKQD-KPVDNAKCSITKGSFRCLSSQFRNETGLITPNKHGLLQNPYRIPPLFIKSPVK 720
QY 992 TKCKKNLLEENFEHSMSPERMGNEI-IPSTVSTISRNIRENVFKKASSNINEVGSS 1050
Db 721 TCKKNVSEKLEHVSPEALGNKSVTQSTVSTISQNNIRENVFKKASSNINEVGSS 774
QY 1051 TNEVGSSINEIGSSDENIQAELGRNRPKLNAMLRLGLVQPEVYKQSLPGSNCKHPEIK 1110
Db 775 -NEVGSSINEIGSSDENIQAELGRNRPKLNAMLRLGLVQPEVYKQSLPGSNCKHPEIK 833
```


QY 1111 QEYEEVQTVNTDPSYLIISDNLEQPMGSSHASQVCSETPDDLLDDGGEIKEDTSFAENDI 1170
| :||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| :
Db 834 QGEKEVYQAYNADEFCLISDNLEQPMGSSCASQVCSETPDDLINDDEIKENISFAESGV 893
QY 1171 KESSAVFSKSVQKGELSRSPFTHTHLAOGYRREGAKKLESSENLSSSE 1219
|| ||||| ||||| : ||||| ||||| : ||||| : ||||| ||||| |||||
Db 894 KERSAVFSKSVQKGKFRRSFPIGRTCLAOGHORRARKLESSENLSSSE 942

Search completed: January 22, 2003, 16:48:27
Job time : 83 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:46:08 ; Search time 18 Seconds
(without alignments)
3045.272 Million cell updates/sec

Title: US-09-734-672-4

Perfect score: 9649

Sequence: 1 MDLSALRVEVQNVINAMOK.....LYQCOELDYLPQIPHSHV 1863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2.6/ptodata/1/1aa/5A_COMB.pep.*

2: /cgn2.6/ptodata/1/1aa/5B_COMB.pep.*

3: /cgn2.6/ptodata/1/1aa/6A_COMB.pep.*

4: /cgn2.6/ptodata/1/1aa/6B_COMB.pep.*

5: /cgn2.6/ptodata/1/1aa/PCTUS_COMB.pep.*

6: /cgn2.6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9649	100.0	1863	1 US-08-798-691-4	Sequence 4, Appli
2	9649	100.0	1863	3 US-08-825-487A-4	Sequence 4, Appli
3	9649	100.0	1863	4 US-09-074-476-6	Sequence 6, Appli
4	9642	99.9	1863	1 US-08-480-784-2	Sequence 2, Appli
5	9642	99.9	1863	1 US-08-483-553-2	Sequence 2, Appli
6	9642	99.9	1863	1 US-08-487-002-2	Sequence 2, Appli
7	9642	99.9	1863	1 US-08-483-554B-2	Sequence 2, Appli
8	9642	99.9	1863	1 US-08-488-011B-2	Sequence 2, Appli
9	9642	99.9	1863	1 US-08-850-727-2	Sequence 2, Appli
10	9642	99.9	1863	5 PCT-US95-10202-2	Sequence 2, Appli
11	9642	99.9	1863	5 PCT-US95-10203-2	Sequence 2, Appli
12	9642	99.9	1863	5 PCT-US95-10220-2	Sequence 2, Appli
13	9635	99.9	1863	1 US-08-598-591-2	Sequence 2, Appli
14	9635	99.9	1863	1 US-08-798-691-2	Sequence 2, Appli
15	9635	99.9	1863	1 US-08-825-487A-6	Sequence 2, Appli
16	9635	99.9	1863	3 US-08-825-487A-2	Sequence 2, Appli
17	9635	99.9	1863	3 US-08-825-487A-6	Sequence 2, Appli
18	9635	99.9	1863	4 US-09-074-476-2	Sequence 2, Appli
19	9635	99.9	1863	4 US-08-425-061-4	Sequence 4, Appli
20	9630	99.8	1863	1 US-08-825-886-16	Sequence 16, Appl
21	9630	99.8	1863	2 US-08-825-886-16	Sequence 16, Appl
22	9581	99.3	1863	2 US-08-603-753D-2	Sequence 2, Appli
23	9581	99.3	1863	4 US-09-099-753-2	Sequence 2, Appli
24	9581	99.3	1863	4 US-08-986-106-2	Sequence 2, Appli
25	9581	99.3	1863	4 US-09-007-678B-49	Sequence 49, Appl
26	9577	99.3	1852	1 US-08-425-061-24	Sequence 24, Appl
27	9577	99.3	1852	2 US-08-825-886-24	Sequence 24, Appl

28	6997	72.5	1363	1 US-08-425-061-23	Sequence 23, Appl
29	6997	72.5	1363	2 US-08-825-886-23	Sequence 23, Appl
30	6222	64.5	1202	1 US-08-425-061-22	Sequence 22, Appl
31	6222	64.5	1202	2 US-08-825-886-22	Sequence 22, Appl
32	4730	49.0	914	1 US-08-425-061-21	Sequence 21, Appl
33	4730	49.0	914	2 US-08-825-886-21	Sequence 21, Appl
34	4621	47.9	900	1 US-08-425-061-20	Sequence 20, Appl
35	4621	47.9	900	2 US-08-825-886-20	Sequence 20, Appl
36	3958	41.0	765	1 US-08-425-061-19	Sequence 19, Appl
37	3958	41.0	765	2 US-08-825-886-19	Sequence 19, Appl
38	1596	16.5	312	1 US-08-425-061-18	Sequence 18, Appl
39	1596	16.5	312	2 US-08-825-886-18	Sequence 18, Appl
40	1563	16.2	309	2 US-08-785-464-5	Sequence 5, Appli
41	1552	16.1	301	2 US-08-785-464-1	Sequence 3, Appli
42	1020	10.6	196	2 US-08-785-464-3	Sequence 2, Appli
43	558	5.8	106	4 US-08-785-464-2	Sequence 3, Appli
44	520	5.4	100	4 US-09-230-196-3	Sequence 4, Appli
45	492	5.1	92	2 US-08-785-464-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-798-691-4
; Sequence 4, Application US/08798691
; Patent No. 5750400
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONCORMED
; STREET: 200 Perry Parkway
; CITY: Gaithersburg
; STATE: MD
; COUNTRY: USA
; ZIP: 20877
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,691
; FILING DATE: 12-Feb-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas Gallegos
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: PA-0054CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-527-2051
; TELEFAX: 301-208-6997
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17

; MAP POSITION: 17q21									
US-08-798-691-4									
Query Match		100.08; Score 9649;		DB 1; Length 1863;					
Best Local Similarity		100.08; Pred. No. 0;							
Matches 1863; Conservative 0;		Mismatches 0;		Indels 0;		Gaps 0;			
QY	1	MDLSALRVEEVQNVINAMQKILECPICLLEIKPEVSTKCDHIFCKFCMLKLLNQKGPQ	60						
DB	1	MDLSALRVEEVQNVINAMQKILECPICLLEIKPEVSTKCDHIFCKFCMLKLLNQKGPQ	60						
QY	61	CPCKNDITKRSLOSTRFPSQVLVBELLKICAFQDQDTGLEVANSYNFAKKNNSPEHLKD	120						
DB	61	CPCKNDITKRSLOSTRFPSQVLVBELLKICAFQDQDTGLEVANSYNFAKKNNSPEHLKD	120						
QY	121	EYSIIQSMGYRNRKRLLOSEPENSLOFTSLSVQLSNGLTVRTLRKQRTQPKTSYVI	180						
DB	121	EYSIIQSMGYRNRKRLLOSEPENSLOFTSLSVQLSNGLTVRTLRKQRTQPKTSYVI	180						
QY	181	ELGSSOSEDVANKATYCSVGDOELLQITPQGTTRDEISLDSAKAACSEFSDVTNTEHHQ	240						
DB	181	ELGSSOSEDVANKATYCSVGDOELLQITPQGTTRDEISLDSAKAACSEFSDVTNTEHHQ	240						
QY	241	PSNNDLNTTEKRAARHPKPEKYQSSVSNLHVPCGCTNTHASSLQHENSSLLLTKDRMNV	300						
DB	241	PSNNDLNTTEKRAARHPKPEKYQSSVSNLHVPCGCTNTHASSLQHENSSLLLTKDRMNV	300						
QY	301	KAFCNKSQOPGLARSOHNRWAGSKETCNDRTPTSTKVKDLNADPLCERKEWNKQLPC	360						
DB	301	KAFCNKSQOPGLARSOHNRWAGSKETCNDRTPTSTKVKDLNADPLCERKEWNKQLPC	360						
QY	361	SENPRDTEVPWITLNSSTIQKWNFWSRDELLGSDSDHGESRNASKAVADVLDVNEVD	420						
DB	361	SENPRDTEVPWITLNSSTIQKWNFWSRDELLGSDSDHGESRNASKAVADVLDVNEVD	420						
QY	421	EYSGSSEKIDLLASDPHALICKSERVSKSVESNIEDKIFGKTYRKKASLPNLSHVTE	480						
DB	421	EYSGSSEKIDLLASDPHALICKSERVSKSVESNIEDKIFGKTYRKKASLPNLSHVTE	480						
QY	481	LIIGAFVTEPQIIQBRPTNLKLRKRRPTSGLHPEDFTKKADLAVOKTPMINOGTNOTE	540						
DB	481	LIIGAFVTEPQIIQBRPTNLKLRKRRPTSGLHPEDFTKKADLAVOKTPMINOGTNOTE	540						
QY	541	QNGQVNIITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPITSSISNMELELNI	600						
DB	541	QNGQVNIITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPITSSISNMELELNI	600						
QY	601	HNSKAPKNRLRRKSTRIHALELVVSRNLSPPNCTELQIDSCSSSEBEIKKKYNOMPV	660						
DB	601	HNSKAPKNRLRRKSTRIHALELVVSRNLSPPNCTELQIDSCSSSEBEIKKKYNOMPV	660						
QY	661	RHSRLQMEGKEPATGAKSKNKPEQTSKRHSDTTPPELKLITNAPGSFTKCSNTSELKE	720						
DB	661	RHSRLQMEGKEPATGAKSKNKPEQTSKRHSDTTPPELKLITNAPGSFTKCSNTSELKE	720						
QY	721	FYNPSLPREEKEKETVKVSNNAEDPKDLMSGERVLQTERSVESSISLVPGCTDYGTO	780						
DB	721	FYNPSLPREEKEKETVKVSNNAEDPKDLMSGERVLQTERSVESSISLVPGCTDYGTO	780						
QY	781	ESISLLEVSTLGAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDEGFKYPLGHEYNHS	840						
DB	781	ESISLLEVSTLGAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDEGFKYPLGHEYNHS	840						
QY	841	RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNGNAEECATFSAHSGSLKKQSPKVT	900						
DB	841	RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNGNAEECATFSAHSGSLKKQSPKVT	900						
QY	901	FECEQKEENQGNESNIKPVOTVNITAGFPVVQKDKPVDNAKSIKGGSRFCLSSQFRG	960						
DB	901	FECEQKEENQGNESNIKPVOTVNITAGFPVVQKDKPVDNAKSIKGGSRFCLSSQFRG	960						
QY	961	NETGLITPNKHGLQNPYRIPPLFPKISFVKTKCKKNLLEENFEHSMSPEREMGNENIP	1020						
DB	961	NETGLITPNKHGLQNPYRIPPLFPKISFVKTKCKKNLLEENFEHSMSPEREMGNENIP	1020						

RESULT 2

US-08-825-487A-4
; Sequence 4, Application US/08825487A
; Patent No. 6048689
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: White, Marga B.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING VARIATIONS IN POLYNUCLEOTIDE SEQUE
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Howrey & Simon
STREET: 1299 Pennsylvania Avenue., N.W.
CITY: Washington,
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/825,487A
FILING DATE: 28-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US98/060002
FILING DATE: 26-Mar-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Albert P. Halluin
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 05371.0012.999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-463-8100
TELEFAX: 650-463-8400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCA1
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
US-08-825-487A-4

Query Match 100.0%; Score 9649; DB 3; Length 1863;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLSALRVEVQVNVINAMQKILECPICLLEIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ 60
DB 1 MDLSALRVEVQVNVINAMQKILECPICLLEIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ 60

QY 61 CPLCKNDITKRSLOESTRFSQVVEELKIIICAFQDGTGLEIYANSYNFAKKNNSPEHLKD 120
DB 61 CPLCKNDITKRSLOESTRFSQVVEELKIIICAFQDGTGLEIYANSYNFAKKNNSPEHLKD 120

QY 121 EVSIIQSMGYRNRAKRLLOSEPENPSLOETSLSVQLSNLGTVRLTKQRIQPKTSVYI 180
DB 121 EVSIIQSMGYRNRAKRLLOSEPENPSLOETSLSVQLSNLGTVRLTKQRIQPKTSVYI 180

QY 181 ELGSDSSEDTVKNATYCSVGDQELLQITPQGTREISLDSAKAACFSETDVTNTEHQ 240
DB 181 ELGSDSSEDTVKNATYCSVGDQELLQITPQGTREISLDSAKAACFSETDVTNTEHQ 240

QY 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVPCGNTNTHASSLQHENSILLTKDRMNYE 300
DB 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVPCGNTNTHASSLQHENSILLTKDRMNYE 300

QY 301 KAEFCNKSQOPGLARSOHRNAGSKETCNDRRTPSTPEKKVVDLADPLCERKENNOKLPC 360
DB 301 KAEFCNKSQOPGLARSOHRNAGSKETCNDRRTPSTPEKKVVDLADPLCERKENNOKLPC 360

QY 361 SENPROTEDPWITLNSSTOKVNEWFSRDELIGSDSDHGESESNKAVADVLVDLYNEVD 420
DB 361 SENPROTEDPWITLNSSTOKVNEWFSRDELIGSDSDHGESESNKAVADVLVDLYNEVD 420

QY 421 EYSGSSEKIDLLASDPHEALICKSERVSHKSVESNIEDIKFGKTYRKKASLPNLSHV TEN 480
DB 421 EYSGSSEKIDLLASDPHEALICKSERVSHKSVESNIEDIKFGKTYRKKASLPNLSHV TEN 480

QY 481 LIIGAFTVTEPQIIQERPLTNKLRKRPTSGHLHPEDFIKADLAVOKTPEMNOGNTQTE 540
DB 481 LIIGAFTVTEPQIIQERPLTNKLRKRPTSGHLHPEDFIKADLAVOKTPEMNOGNTQTE 540

QY 541 QNGOVNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI 600
DB 541 QNGOVNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI 600

QY 601 HNSKAPKNRLRRKSSRTHIALELVVSRNLSPNCTELQIDSCSSSEIEIKKKYNQMPV 660
DB 601 HNSKAPKNRLRRKSSRTHIALELVVSRNLSPNCTELQIDSCSSSEIEIKKKYNQMPV 660

QY 661 RHNRNLQMEGKEPATGAKKSNKNPNQOTSKRHDSOTFPELKLTNAPGSKTCSNTSELKE 720
DB 661 RHNRNLQMEGKEPATGAKKSNKNPNQOTSKRHDSOTFPELKLTNAPGSKTCSNTSELKE 720

QY 721 FVNPSLPREEKEKLETVKVSNNADPKDMLSGERVLOTERSVESSSISLVPGTDYGTQ 780
DB 721 FVNPSLPREEKEKLETVKVSNNADPKDMLSGERVLOTERSVESSSISLVPGTDYGTQ 780

QY 781 ESISLLEVSTLKGAKTEPNKCVSQCAAFENPKGLIHGCSKDRNDTEGFKYPLGHEVNH 840
DB 781 ESISLLEVSTLKGAKTEPNKCVSQCAAFENPKGLIHGCSKDRNDTEGFKYPLGHEVNH 840

QY 841 RETSIEMESELDAQYLQNTFKVSRQSFALPSNPGNAEEECATFSAHSGSLKKQSPKVT 900
DB 841 RETSIEMESELDAQYLQNTFKVSRQSFALPSNPGNAEEECATFSAHSGSLKKQSPKVT 900

QY 901 FECEQKEENOGKNESNIKPVQTVNITAGFPVVGOKDPVDNAKCSIKGGSRCFLSSQPRG 960
DB 901 FECEQKEENOGKNESNIKPVQTVNITAGFPVVGOKDPVDNAKCSIKGGSRCFLSSQPRG 960

QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNLLNEENFEEHSMSPEREMGNENIP 1020
DB 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNLLNEENFEEHSMSPEREMGNENIP 1020

QY 1021 STVSTISRNIRNENVEKEASSNINEVSGSTNEVGSSINEIGSSDENIOAELGRNRPKL 1080
DB 1021 STVSTISRNIRNENVEKEASSNINEVSGSTNEVGSSINEIGSSDENIOAELGRNRPKL 1080

QY 1081 NAMRLGVLPQPEYVKOSLPGSNCKHPEIKKOEYEEVVQVNTDFSPYLLSDNLEOPMGSS 1140
DB 1081 NAMRLGVLPQPEYVKOSLPGSNCKHPEIKKOEYEEVVQVNTDFSPYLLSDNLEOPMGSS 1140

QY 1141 HASQVCSFETPDLLDDGEIKEDTSPAENDIKESSAVFSSQVKGELSRSPSPFTHLQAQ 1200
DB 1141 HASQVCSFETPDLLDDGEIKEDTSPAENDIKESSAVFSSQVKGELSRSPSPFTHLQAQ 1200

QY 1201 GYRRGAKKLESSEENLSSDEELPCFQHLFGKVNNIPSOQSTRHSTVATECLSKTEENL 1260
DB 1201 GYRRGAKKLESSEENLSSDEELPCFQHLFGKVNNIPSOQSTRHSTVATECLSKTEENL 1260

QY 1261 LSLKNSLDCNQVILAKASQEHHLSEETKCSASLFSOCSSELEDLTANTNTQDPFLIGS 1320
DB 1261 LSLKNSLDCNQVILAKASQEHHLSEETKCSASLFSOCSSELEDLTANTNTQDPFLIGS 1320

QY 1321 SKOMRHQSESQGVGLSDKELVSDDEERGTLLENQEQSMDSNLGEAASGCESETSVSE 1380
DB 1321 SKOMRHQSESQGVGLSDKELVSDDEERGTLLENQEQSMDSNLGEAASGCESETSVSE 1380

QY 1381 DCSGLSSQSDIITTOQRTMQHNLKLOQEMAELEAVLBOHGSQPSNSYPSIISOSSALE 1440
DB 1381 DCSGLSSQSDIITTOQRTMQHNLKLOQEMAELEAVLBOHGSQPSNSYPSIISOSSALE 1440

QY 1441 DLNRNPEQSTSEKAVLTSSQSEVPISQNPGLSADKFEVSADSSSTSKNKEPVERSSPSK 1500
DB 1441 DLNRNPEQSTSEKAVLTSSQSEVPISQNPGLSADKFEVSADSSSTSKNKEPVERSSPSK 1500

QY 1501 CPSLDDRWYMHSCSGSLQNRNYPQBELIKVVDVBEQOLESSESGPHDLTETSYLPRQDLEG 1560

Db 1501 CFSLDRWYMHSCSGSLQNRNYPQSEELIKVVDVEEQLEESGPHDLTETSYLRQDLEG 1560
QY 1561 TPYLSGSLFSDPESDPSDRAPESARVGNIPSSTSALKVPQLKVAESAQSPAAHTT 1620
Db 1561 TPYLSGSLFSDPESDPSDRAPEASRVGNIPSSTSALKVPQLKVAESAQSPAAHTT 1620
QY 1621 DTAGYNAMESYSREKPELTASTERVNKRMSMVVSGLTPEEFMLVYKFAKHHITLTNLI 1680
Db 1621 DTAGYNAMESYSREKPELTASTERVNKRMSMVVSGLTPEEFMLVYKFAKHHITLTNLI 1680
QY 1681 TEETHVVMKTDAEFVCERTLKYFLGIAGGKVVSYFWVTQSIKERKMLNHDHFEVRGDV 1740
Db 1681 TEETHVVMKTDAEFVCERTLKYFLGIAGGKVVSYFWVTQSIKERKMLNHDHFEVRGDV 1740
QY 1741 VNGRNHQPKRARESDQRKIFRGLIEICYGPTNNPTDQLEWVQVLCASVVKELSSFTL 1800
Db 1741 VNGRNHQPKRARESDQRKIFRGLIEICYGPTNNPTDQLEWVQVLCASVVKELSSFTL 1800
QY 1801 GTGVHPVIVVQPDAMTONGFHAIGOMCEAPVVTREWVLDVALYQCOELDTYLIPOIPH 1860
Db 1801 GTGVHPVIVVQPDAMTONGFHAIGOMCEAPVVTREWVLDVALYQCOELDTYLIPOIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 3
US-09-074-476-6
; Sequence 6, Application US/09074476
; Patent No. 6130322
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Thurber, Denise
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; TITLE OF INVENTION: BRCA1 Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N. W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,476
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/074,453
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Albert P. Halluin
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: 5371.34.US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid

; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1 (Om13)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
US-09-074-476-6

Query Match 100.0%; Score 9649; DB 4; Length 1863;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLSARVEEVQNVINAMQKILECPICILELIKEPVSTKCDHFCKFCMLKLLNKKGPSQ 60
Db 1 MDLSARVEEVQNVINAMQKILECPICILELIKEPVSTKCDHFCKFCMLKLLNKKGPSQ 60
QY 61 CPLCKNDITKRSLOESTRFVSOLVVEELLKIIICAFQDTCLEAYANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFVSOLVVEELLKIIICAFQDTCLEAYANSYNFAKKENNSPEHLKD 120
QY 121 EVSIIQSMGYNRNRAKRLLOSEPENPSLQETSLSVQLSNLGTVRTLRKORIQPOKTSVYI 180
Db 121 EVSIIQSMGYNRNRAKRLLOSEPENPSLQETSLSVQLSNLGTVRTLRKORIQPOKTSVYI 180
QY 181 ELGSDSSEDVTNKATYCSVGQDQLLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDVTNKATYCSVGQDQLLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERPEKYQGVSSVNLHVPCGTNTHASSLOHENSLLLTDKDMNVE 300
Db 241 PSNNDLNTTEKRAAERPEKYQGVSSVNLHVPCGTNTHASSLOHENSLLLTDKDMNVE 300
QY 301 KAEECNKSKOPGLARSQHRNWAWSKETCNDRTPTSTKVKVDLADPLCERKEWNKQKLPC 360
Db 301 KAEECNKSKOPGLARSQHRNWAWSKETCNDRTPTSTKVKVDLADPLCERKEWNKQKLPC 360
QY 361 SENPRDTEVPWITLNSIOKVNEWFSRDELLSGSDSDHGESESNKAVADVLOVLNEVD 420
Db 361 SENPRDTEVPWITLNSIOKVNEWFSRDELLSGSDSDHGESESNKAVADVLOVLNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTE 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTE 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINOQTNQTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINOQTNQTE 540
QY 541 QNGOVNMTNSGHENKTKGDSIQNEKNPNPTESLEKSAFKAETPSSISNNELELNI 600
Db 541 QNGOVNMTNSGHENKTKGDSIQNEKNPNPTESLEKSAFKAETPSSISNNELELNI 600
QY 601 HNSKAPKNRLRRKSTRHITHALELVVSRNLSPPNCTELOIDSCSSSDEIKKKYNQMPV 660
Db 601 HNSKAPKNRLRRKSTRHITHALELVVSRNLSPPNCTELOIDSCSSSDEIKKKYNQMPV 660
QY 661 RHSRNLQMEGKEPATGAKKSNKPNEQTSKRHSDTPELKLITNAPGSFTKCSNTSELKE 720
Db 661 RHSRNLQMEGKEPATGAKKSNKPNEQTSKRHSDTPELKLITNAPGSFTKCSNTSELKE 720
QY 721 FVNPSPREEKEKLETYKVSNNNAEDPKDMLSGERVLTQERSVESSSISLVPQTDYGTQ 780
Db 721 FVNPSPREEKEKLETYKVSNNNAEDPKDMLSGERVLTQERSVESSSISLVPQTDYGTQ 780
QY 781 ESISLLEVSITLGRKAKTEPNKCVSQAFAENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
Db 781 ESISLLEVSITLGRKAKTEPNKCVSQAFAENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLQNTFKVSKRQSFAFNSPNAEECAFTSAHSGSLKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSKRQSFAFNSPNAEECAFTSAHSGSLKQSPKVT 900

Db 841 RETSIEMBESELDQYQLQNTFKYKRSQSFALFSPNGNAEBECATFSAHSGSLKKQSPKVT 900
QY 901 FECQKEENQKGNESNKPQVTNITAGFPVVGQKDPVDNAKCSIKGGSRFCLSSQFRG 960
Db 901 FECQKEENQKGNESNKPQVTNITAGFPVVGQKDPVDNAKCSIKGGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNYRIPPLPIKSFVKTKCKNLLBENFEHSHMSPEREMGNINIP 1020
Db 961 NETGLITPNKHGLLQNYRIPPLPIKSFVKTKCKNLLBENFEHSHMSPEREMGNINIP 1020
QY 1021 STVSTISRNIRNVKFASSNINEVSGSTNEVSGSINEIGSSDENIQALGRNRGPKL 1080
Db 1021 STVSTISRNIRNVKFASSNINEVSGSTNEVSGSINEIGSSDENIQALGRNRGPKL 1080
QY 1081 NAMLRGLVQPEYKYQLPGSNCKHPKIKQYEEVVQTVNTDFSPYLIISDNLEQPMGSS 1140
Db 1081 NAMLRGLVQPEYKYQLPGSNCKHPKIKQYEEVVQTVNTDFSPYLIISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDDLLDDGEIKEDTSAFANDIKESSAVFSKVQKGEJSRSPSPHTHQAQ 1200
Db 1141 HASQVCSETPDDLLDDGEIKEDTSAFANDIKESSAVFSKVQKGEJSRSPSPHTHQAQ 1200
QY 1201 GYRRGAKKLESSSENLSSEDEELPCFOHLLFGKVNNTPSQSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSSENLSSEDEELPCFOHLLFGKVNNTPSQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNOVILAKASQEHLSBETKCSASLSFSSQSELEDLTANTNTODPFLIGS 1320
Db 1261 LSLKNSLNDCSNOVILAKASQEHLSBETKCSASLSFSSQSELEDLTANTNTODPFLIGS 1320
QY 1321 SKOMRHOSESQVGLSKDELVSDDERGTGLEENNOBEQSMDSNLGPAASGCESETSVSE 1380
Db 1321 SKOMRHOSESQVGLSKDELVSDDERGTGLEENNOBEQSMDSNLGPAASGCESETSVSE 1380
QY 1381 DCSGLSSQSDILITQOQDRTWOHNLKIQOEMAELEAVLEOHGQSPNSYPSIISDSALE 1440
Db 1381 DCSGLSSQSDILITQOQDRTWOHNLKIQOEMAELEAVLEOHGQSPNSYPSIISDSALE 1440
QY 1441 DLNRPEQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPVGVERSPSK 1500
Db 1441 DLNRPEQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPVGVERSPSK 1500
QY 1501 CPSLDDRMYHSCSGSLQNRNYPQSEBELIKVVDVEEQOLEESGPHDLTETSYLPRODLEG 1560
Db 1501 CPSLDDRMYHSCSGSLQNRNYPQSEBELIKVVDVEEQOLEESGPHDLTETSYLPRODLEG 1560
QY 1561 TPYLESGISLFSDDPESDPSDRAPESARYGNIPSPSTTSALKVPOLKVAESAQSPAHAHT 1620
Db 1561 TPYLESGISLFSDDPESDPSDRAPESARYGNIPSPSTTSALKVPOLKVAESAQSPAHAHT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERNKRMVMVYVGLTPEEFMLVYKFARKHHTLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERNKRMVMVYVGLTPEEFMLVYKFARKHHTLTNLI 1680
QY 1681 TEETHVVMKTDAEFVCERTLKVFLGIAGGKWVYVFWVTQSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDAEFVCERTLKVFLGIAGGKWVYVFWVTQSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRHHGPKRARESQRDKLFRGLEICCCYGFPTNMPDQLEWVQLCGASVVKELSSFTL 1800
Db 1741 VNGRHHGPKRARESQRDKLFRGLEICCCYGFPTNMPDQLEWVQLCGASVVKELSSFTL 1800
QY 1801 GTGVHPPIVVVQPDWATBDNGFHAIGOMCEAPVVTREWLDSVALYQCOELDTYLIPQIPH 1860
Db 1801 GTGVHPPIVVVQPDWATBDNGFHAIGOMCEAPVVTREWLDSVALYQCOELDTYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 4

US-08-480-784-2

; Sequence 2, Application US/08480784
; Patent No. 5693473
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Beetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,784
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-784-2

Query Match 99.9%; Score 9642; DB 1; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQVNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ 60
Db 1 MDLSALRVEEVQVNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ 60
QY 61 CPLCKNDITKRSLOESTRFSQVLSQVEELLKTIKICAFQDGTGLGYANSYNFACKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVLSQVLEELLKTIKICAFQDGTGLGYANSYNFACKENNSPEHLKD 120

QY 121 EVSTIOSMGYNRAKRLLOSEPENPSLOETSLVSQLSNLTGVRTLRTKORIQOQKTSYVI 180
DB 121 EVSLIOSMGIRNAKRLLOSEPENPSLOETSLVSQLSNLTGVRTLRTKORIQOQKTSYVI 180
QY 181 ELGSDSSEDTVNKATYCSVGDDQLQITPQGRTRDEISLSDSAKKAACEFSETDVTNTEHHQ 240
DB 181 ELGSDSSEDTVNKATYCSVGDDQLQITPQGRTRDEISLSDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGCTNTHASSLOHENSLLLTUKORMNVE 300
DB 241 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGCTNTHASSLOHENSLLLTUKORMNVE 300
QY 301 KAEFCNKSQKPLARSOHNRWAGSKETCNDRTPTSTPEKKVVDLADPLCKERKWNKQKLP 360
DB 301 KAEFCNKSQKPLARSOHNRWAGSKETCNDRTPTSTPEKKVVDLADPLCKERKWNKQKLP 360
QY 361 SENPRDTEVPWITLNSSIOKWNEMFSRSDLLGSDSDHSGESENNAKVADVDLNEVD 420
DB 361 SENPRDTEVPWITLNSSIOKWNEMFSRSDLLGSDSDHSGESENNAKVADVDLNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHKSYESNIEDKIFGKTYRKKAASLPNLSHVTEN 480
DB 421 EYSGSSEKIDLLASDPHEALICKSERVHKSYESNIEDKIFGKTYRKKAASLPNLSHVTEN 480
QY 481 LIIGAFVTEPQIIQERPLNKLKRRRPTSGLHPEDFTKKADLAVOKTPEMINOQTNTE 540
DB 481 LIIGAFVTEPQIIQERPLNKLKRRRPTSGLHPEDFTKKADLAVOKTPEMINOQTNTE 540
QY 541 QNGQVNNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELENI 600
DB 541 QNGQVNNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELENI 600
QY 601 HNSKAPKNRLRKSSTRIHHALELVVSNLSPNCTELOIDSCSSSEIEKKKYNQMPV 660
DB 601 HNSKAPKNRLRKSSTRIHHALELVVSNLSPNCTELOIDSCSSSEIEKKKYNQMPV 660
QY 661 RHSRNLQMEGEPATGAKSKNPNQTSKRHSDTPELKLITNAPGFTKCSNTSELKE 720
DB 661 RHSRNLQMEGEPATGAKSKNPNQTSKRHSDTPELKLITNAPGFTKCSNTSELKE 720
QY 721 FVNPSLPREKEBEKLETVKVSNAEDPKDMLSGSERVLTQERSVSSSISLVPGDYGTQ 780
DB 721 FVNPSLPREKEBEKLETVKVSNAEDPKDMLSGSERVLTQERSVSSSISLVPGDYGTQ 780
QY 781 ESTSLLEVSTLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
DB 781 ESTSLLEVSTLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMERSELDQYLOQNTFKVSKRQSPALFSPNGNAEECATPSAHSGLSKQSPKVT 900
DB 841 RETSIEMERSELDQYLOQNTFKVSKRQSPALFSPNGNAEECATPSAHSGLSKQSPKVT 900
QY 901 FECEQEENOGKNESNIKPVQNTVNTAGPVVQKDKPVDNAKCSIKGGRFCLSSQFRG 960
DB 901 FECEQEENOGKNESNIKPVQNTVNTAGPVVQKDKPVDNAKCSIKGGRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLPIKFSVTKCKKNLLEENFEHSMSPREMGNEINIP 1020
DB 961 NETGLITPNKHGLLQNPYRIPPLPIKFSVTKCKKNLLEENFEHSMSPREMGNEINIP 1020
QY 1021 STVSTISRNNIRENVFKEASSNINEVSSSTNEVGSSINEIGSSDENQAEILGRNRGPKL 1080
DB 1021 STVSTISRNNIRENVFKEASSNINEVSSSTNEVGSSINEIGSSDENQAEILGRNRGPKL 1080
QY 1081 NAMRLGVLOPEVYKQSLPGSNCKHPEIKKQBYEEVQTVNTDPSYLLISDNLEQPMGSS 1140
DB 1081 NAMRLGVLOPEVYKQSLPGSNCKHPEIKKQBYEEVQTVNTDPSYLLISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDOLLDDGEIKEDTSAENDIKESSAVFSKSVQKGLSRSPFFTHLQ 1200
DB 1141 HASQVCSETPDOLLDDGEIKEDTSAENDIKESSAVFSKSVQKGLSRSPFFTHLQ 1200

QY 1201 GYRRGAKKLESSEENLSEDELPQFQHLFGKVNIPISQSTRHSTVATECLSKNTEENL 1260
DB 1201 GYRRGAKKLESSEENLSEDELPQFQHLFGKVNIPISQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDSCNOVILAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPFLIGS 1320
DB 1261 LSLKNSLNDSCNOVILAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHQSESOGVGLSKDELVSDDERGTGLBENNQEQSMDSNLGEAASGESETSVSE 1380
DB 1321 SKQMRHQSESOGVGLSKDELVSDDERGTGLBENNQEQSMDSNLGEAASGESETSVSE 1380
QY 1381 DCSGLSSOSDILTTQORDTMQHNLIKLOQEMAELEAVLEQHGQSQPSNSYPSIISDSSALE 1440
DB 1381 DCSGLSSOSDILTTQORDTMQHNLIKLOQEMAELEAVLEQHGQSQPSNSYPSIISDSSALE 1440
QY 1441 DLRNPEQSTSEKAVLTSQKSSEYPIQSONPEGLSADKFEVSADSSTSKNKEPGEVERSSPSK 1500
DB 1441 DLRNPEQSTSEKAVLTSQKSSEYPIQSONPEGLSADKFEVSADSSTSKNKEPGEVERSSPSK 1500
QY 1501 CPSLDDRWYMHSCSGLQNRNYPQOEELIKVYDVEEQLEESGPHDLTETSYPQLQDLEG 1560
DB 1501 CPSLDDRWYMHSCSGLQNRNYPQOEELIKVYDVEEQLEESGPHDLTETSYPQLQDLEG 1560
QY 1561 TPYLESGLISLFSDDPESDPSEDRAPEARVGNIPSTTSALKVPQLKVAESAQSPAAAHTT 1620
DB 1561 TPYLESGLISLFSDDPESDPSEDRAPEARVGNIPSTTSALKVPQLKVAESAQSPAAAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSMVVSGLTPEEPMLVYKFARKHHTLTNLI 1680
DB 1621 DTAGYNAMESVSREKPELTASTERVNKRMSMVVSGLTPEEPMLVYKFARKHHTLTNLI 1680
QY 1681 TEETHVVMKTDAEVCEBTLKYLFIAGGKVVVYFWVTQSIKERKMLNEHDFEVRGDV 1740
DB 1681 TEETHVVMKTDAEVCEBTLKYLFIAGGKVVVYFWVTQSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQGPKRARESDQRKIFRGLIEICCYGPTNPTDQLEMMVOLCGASVVKELSSFTL 1800
DB 1741 VNGRNHQGPKRARESDQRKIFRGLIEICCYGPTNPTDQLEMMVOLCGASVVKELSSFTL 1800
QY 1801 GTGVHPIVVQPDAMTENDNGFHAIGQMCEAPVVTREWVLDVSVALYQCELDYLIPOIPH 1860
DB 1801 GTGVHPIVVQPDAMTENDNGFHAIGQMCEAPVVTREWVLDVSVALYQCELDYLIPOIPH 1860
QY 1861 SHY 1863
DB 1861 SHY 1863

RESULT 5

US-08-483-553-2
; Sequence 2, Application US/08483553
; Patent No. 5709999
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA

; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,553
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-483-553-2

Query Match 99.9%; Score 9642; DB 1; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQVNVINAMQKILEPICLELIKBPVSTKCDHIFCKFCMLKLLNKKGPSQ 60
Db 1 MDLSALRVEEVQVNVINAMQKILEPICLELIKBPVSTKCDHIFCKFCMLKLLNKKGPSQ 60

QY 61 CPLCKNDITKRSIQESTRFSQVQDQLLQITPGTRDEISLSAKKAACEFSETDVTNTEHHQ 240
Db 61 CPLCKNDITKRSIQESTRFSQVQDQLLQITPGTRDEISLSAKKAACEFSETDVTNTEHHQ 240

QY 121 EVSIQSMGYRNRKRLQISEPENPSLOETSLSVOLSNLGTVRTLTKQRIQOKT SVYI 180
Db 121 EVSIQSMGYRNRKRLQISEPENPSLOETSLSVOLSNLGTVRTLTKQRIQOKT SVYI 180

QY 181 ELGSDSSEDVTNKTCSVGDQELLQITPGTRDEISLSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDVTNKTCSVGDQELLQITPGTRDEISLSAKKAACEFSETDVTNTEHHQ 240

QY 241 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGNTNTHASSLQHENSLLLT KDRMNYE 300
Db 241 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGNTNTHASSLQHENSLLLT KDRMNYE 300

QY 301 KAEFCNKSQKQGLARQHNRWAGSKETCNDNRRTPSTPEKKVVDLNADPLCERKENNKKLPC 360
Db 301 KAEFCNKSQKQGLARQHNRWAGSKETCNDNRRTPSTPEKKVVDLNADPLCERKENNKKLPC 360

QY 361 SENPRTEDEVPTITLNSSTQKVNWF SRSDLLGSDSDSHDGSESNAKVADVLDVLNEVD 420
Db 361 SENPRTEDEVPTITLNSSTQKVNWF SRSDLLGSDSDSHDGSESNAKVADVLDVLNEVD 420

QY 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTE 480

Db 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTE 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKKADLAVOKTPEMINQGNQTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKKADLAVOKTPEMINQGNQTE 540
QY 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMWELENI 600
Db 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMWELENI 600
QY 601 HNSKAPKKNRLRRKSSSTRHIALELVVSRNLSPNCTELQIDSCSSSEIEKKKKYNOMPV 660
Db 601 HNSKAPKKNRLRRKSSSTRHIALELVVSRNLSPNCTELQIDSCSSSEIEKKKKYNOMPV 660
QY 661 RHSRNLQLMGKREPATGAKKSNKPNNEOTSKRHDSDFPELKLTNAPGSTKCSNTSELKE 720
Db 661 RHSRNLQLMGKREPATGAKKSNKPNNEOTSKRHDSDFPELKLTNAPGSTKCSNTSELKE 720
QY 721 FVNPSLPREEKEEKLTVKVSNNADPDKDMLSGERVLOTERRSVESSSISLVPGTDTGQT 780
Db 721 FVNPSLPREEKEEKLTVKVSNNADPDKDMLSGERVLOTERRSVESSSISLVPGTDTGQT 780
QY 781 ESISLLEVTSLGKAKTEPNKCVSQAAPENPKGLIHGCKDNRNDTEGFKYPLGHEVNH 840
Db 781 ESISLLEVTSLGKAKTEPNKCVSQAAPENPKGLIHGCKDNRNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDQAOLQNTFKVSKQSFALFSPNPNABEECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDQAOLQNTFKVSKQSFALFSPNPNABEECATFSAHSGSLKKQSPKVT 900
QY 901 FECEQKEENQGNKESNIPVQITVNIITAGVPVQKDPVDNAKCSIKGSRFCLSSQFRG 960
Db 901 FECEQKEENQGNKESNIPVQITVNIITAGVPVQKDPVDNAKCSIKGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLFIKSFVTKCKKNLLEENFEHSMSPEREMGNENIP 1020
Db 961 NETGLITPNKHGLLQNPYRIPPLFIKSFVTKCKKNLLEENFEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIQAEGLRNRGPKL 1080
Db 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIQAEGLRNRGPKL 1080
QY 1081 NAMRLGVLPQPEYVKOSLPGSNCKHPEIKKQEEYEVVQVNTDFSPYILSDNLEQPMGSS 1140
Db 1081 NAMRLGVLPQPEYVKOSLPGSNCKHPEIKKQEEYEVVQVNTDFSPYILSDNLEQPMGSS 1140
QY 1141 HASQVCSETPDDLLDDGEIKEDTSFANDIKESSAVFSKVQKGLSRSPSPFTHHLAQ 1200
Db 1141 HASQVCSETPDDLLDDGEIKEDTSFANDIKESSAVFSKVQKGLSRSPSPFTHHLAQ 1200
QY 1201 GYRRGAKKLESSEENLSSDEELPCFQHLHFGKVNNIPSOSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSSDEELPCFQHLHFGKVNNIPSOSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNQVILAKAQEHLSSEETKCSASFSSOCSELEDLTANTNTQDPLIGS 1320
Db 1261 LSLKNSLNDCSNQVILAKAQEHLSSEETKCSASFSSOCSELEDLTANTNTQDPLIGS 1320
QY 1321 SKOMRHQSESGVGSLDKELVSDDEERGTLGLENQEQSDMDSNLGEEASGCESETSVSE 1380
Db 1321 SKOMRHQSESGVGSLDKELVSDDEERGTLGLENQEQSDMDSNLGEEASGCESETSVSE 1380
QY 1381 DCSGLSSQSDILTTQORDTMQHNLIKLOQEMAELEAVLEHQSGSQSPNSYPSIISSSALE 1440
Db 1381 DCSGLSSQSDILTTQORDTMQHNLIKLOQEMAELEAVLEHQSGSQSPNSYPSIISSSALE 1440
QY 1441 DLRNPQOSTSEKAVLTISQKSSSEYPISONPEGLSADKFYSADSTSKNKEPGVERSSPSK 1500
Db 1441 DLRNPQOSTSEKAVLTISQKSSSEYPISONPEGLSADKFYSADSTSKNKEPGVERSSPSK 1500
QY 1501 CPSSLDRWYMHSCGSLQNRNTPSQBELIKVYVDVEBQQLLEESGPHDLTETSYLPRODLEG 1560

Db 1501 CFSLODRWYMHSCSGSLQNRNYPQOEELIKVVVDVEOQLEESGPHDLTETSYLPQDLEG 1560
Qy 1561 TPYLESGISLFDSPEDPESDRAPESARVGNIPSTTSALKVPOLKVAESAQSAAAHTT 1620
Db 1561 TPYLESGISLFDSPEDPESDRAPESARVGNIPSTTSALKVPOLKVAESAQSAAAHTT 1620
Qy 1621 DTAGYNAMESYSREKPELTASTERVNKRMSMVYSGLTPEEFMLVYFARKHHITLTNLI 1680
Db 1621 DTAGYNAMESYSREKPELTASTERVNKRMSMVYSGLTPEEFMLVYFARKHHITLTNLI 1680
Qy 1681 TEETHVVKTKDAEFVCERTLKYFLGIAGGKWWVSYPVWTOSIKERKMLNHDHFEVRGDV 1740
Db 1681 TEETHVVKTKDAEFVCERTLKYFLGIAGGKWWVSYPVWTOSIKERKMLNHDHFEVRGDV 1740
Qy 1741 VNGRHHQGPKRAESQDRKIRFGLGICCYGPTNNMPTDQLEWMYVOLCGASVVKELSFTL 1800
Db 1741 VNGRHHQGPKRAESQDRKIRFGLGICCYGPTNNMPTDQLEWMYVOLCGASVVKELSFTL 1800
Qy 1801 GTGVHPVVVQPDAWTNGFHAIGQMCEAPVVTREWWLDSVALYQCELDTYLIPQIPH 1860
Db 1801 GTGVHPVVVQPDAWTNGFHAIGQMCEAPVVTREWWLDSVALYQCELDTYLIPQIPH 1860
Qy 1861 SHY 1863
Db 1861 SHY 1863

RESULT 6
US-08-487-002-2
; Sequence 2, Application US/08487002
; Patent No. 5710001
; GENERAL INFORMATION:
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Simard, Jacques
; APPLICANT: Emi, Mitsuru
; APPLICANT: Nakamura, Yusuke
; APPLICANT: Durocher, Francine
; TITLE OF INVENTION: 17q-linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,002
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.

; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-487-002-2

Query Match 99.9%; Score 9642; DB 1; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MDLSARVEEVQNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLNQKKGPSQ 60
Db 1 MDLSARVEEVQNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLNQKKGPSQ 60
Qy 61 CPICKNDITKRSLOESTRFSQVLEELLKIIICAFQLDTGLEYANSYNFAKKENNSPEHLKD 120
Db 61 CPICKNDITKRSLOESTRFSQVLEELLKIIICAFQLDTGLEYANSYNFAKKENNSPEHLKD 120
Qy 121 EVSIIQSMGYRNRAKRLLOSEPENPSLOETSLSVQLSNLGTVRTLRTKQRIPOQKTSYVI 180
Db 121 EVSIIQSMGYRNRAKRLLOSEPENPSLOETSLSVQLSNLGTVRTLRTKQRIPOQKTSYVI 180
Qy 181 ELGSDSSEDVTNKATYCSVGDDQELLQITPQGTTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDVTNKATYCSVGDDQELLQITPQGTTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Qy 241 PSNNDLNTTEKRAAERPEKYQGSVSNLHVEPCGTTNTHASSLOHENSLLLTKDRMNV 300
Db 241 PSNNDLNTTEKRAAERPEKYQGSVSNLHVEPCGTTNTHASSLOHENSLLLTKDRMNV 300
Qy 301 KAEFCNKSKOPGLARSOHNRWAGSKETCNDRTPTSTKPKVDLADPLCERKEWNKQKLPC 360
Db 301 KAEFCNKSKOPGLARSOHNRWAGSKETCNDRTPTSTKPKVDLADPLCERKEWNKQKLPC 360
Qy 361 SENPRDTEVPWITLNSSIQKVNEWFSRSDDELGSDSDHGESESNKAVADVLDVLEVD 420
Db 361 SENPRDTEVPWITLNSSIQKVNEWFSRSDDELGSDSDHGESESNKAVADVLDVLEVD 420
Qy 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVSNEIEDKIFGKTYRKASLPNLSHVTE 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVSNEIEDKIFGKTYRKASLPNLSHVTE 480
Qy 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPDEFKKAADLAVQKTPMINOCTNOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPDEFKKAADLAVQKTPMINOCTNOTE 540
Qy 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTLESLEKESAFKTAEPITSSSISSNMELE 600
Db 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTLESLEKESAFKTAEPITSSSISSNMELE 600
Qy 601 HNSKAPKKNLRLRRKSTRHIALELVYSRNLSPNCTELOIDSCSSSEIEIKKKYQNPV 660
Db 601 HNSKAPKKNLRLRRKSTRHIALELVYSRNLSPNCTELOIDSCSSSEIEIKKKYQNPV 660
Qy 661 RHSRNLQMEGKEPATGAKKSNKNEOTSKRHSDTPELKLITNAPGSFTKCSNTSELKE 720
Db 661 RHSRNLQMEGKEPATGAKKSNKNEOTSKRHSDTPELKLITNAPGSFTKCSNTSELKE 720
Qy 721 FVNPSPREEKEEKLETYKVSNNADPKDMLSGERVLTQTSRVSESSISLVPGTDTYGTQ 780
Db 721 FVNPSPREEKEEKLETYKVSNNADPKDMLSGERVLTQTSRVSESSISLVPGTDTYGTQ 780
Qy 781 ESISLLEVSTLGRKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNH 840
Db 781 ESISLLEVSTLGRKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNH 840

QY 841 RETSIEMESELDAQYLQNTFKVSKROSFALFSPNGNAEEBCATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSKROSFAPFSPNGNAEEBCATFSAHSGSLKKQSPKVT 900
QY 901 FECQKQENQCKNESNKPQVTYNITAGFPVVGOKDPVDNAKCSIKGGSRFCLLSSQFRG 960
Db 901 FECQKQENQCKNESNKPQVTYNITAGFPVVGOKDPVDNAKCSIKGGSRFCLLSSQFRG 960
QY 961 NETGLITPNKHGGLQNPYRPPLEPIKFSFKTKCKKNLLENFEEHSMSPEREMGNENIP 1020
Db 961 NETGLITPNKHGGLQNPYRPPLEPIKFSFKTKCKKNLLENFEEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNINRENVFKEASSNINEYSSSTNEVGSSINEIGSSDENIQAEGLGRNRGPKL 1080
Db 1021 STVSTISRNINRENVFKEASSNINEYSSSTNEVGSSINEIGSSDENIQAEGLGRNRGPKL 1080
QY 1081 NAMLRGLVLOPEVYKQSLPGSNCKHPKIKQYEEVQTVNTDFSPYLISDNLEQPMGSS 1140
Db 1081 NAMLRGLVLOPEVYKQSLPGSNCKHPKIKQYEEVQTVNTDFSPYLISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDDLLDGEIKEDTSAFENDIKESSAVFSKSVOKGELSRSPFTHTHLAQ 1200
Db 1141 HASQVCSETPDDLLDGEIKEDTSAFENDIKESSAVFSKSVOKGELSRSPFTHTHLAQ 1200
QY 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFGKVNNTIPSOSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFGKVNNTIPSOSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLDCSNQVILAKAQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLDCSNQVILAKAQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHOSQGVGLSKELVSDDEERGTGLEENNOEQSMDSNLGEAAGCESETSVSE 1380
Db 1321 SKQMRHOSQGVGLSKELVSDDEERGTGLEENNOEQSMDSNLGEAAGCESETSVSE 1380
QY 1381 DCSGLSSQSDILTTQQRDTMOHNLKIQEAMAEVLEQHGQSPNSYPSIISDSSALE 1440
Db 1381 DCSGLSSQSDILTTQQRDTMOHNLKIQEAMAEVLEQHGQSPNSYPSIISDSSALE 1440
QY 1441 DLNRPEOSTSEKAVLTQKSEYPISONPEGLSADKFEVSADSTSKNKEPGEVRSRSPSK 1500
Db 1441 DLNRPEOSTSEKAVLTQKSEYPISONPEGLSADKFEVSADSTSKNKEPGEVRSRSPSK 1500
QY 1501 CPSLDDRWYHSCSGSLQNRNYPQELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDDRWYHSCSGSLQNRNYPQELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESGISLFSDDPESDRAPEARVGNIPSSTSALKVPOLKVAESAQAASAAHTT 1620
Db 1561 TPYLESGISLFSDDPESDRAPEARVGNIPSSTSALKVPOLKVAESAQAASAAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERNVKNRSMVYSGLTPEEFMLVYKFAKHHITLNL 1680
Db 1621 DTAGYNAMESVSREKPELTASTERNVKNRSMVYSGLTPEEFMLVYKFAKHHITLNL 1680
QY 1681 TEETHVVMKTDABFCERTLKVFLGAGGWVYSYFWWTQSIKERKMLNEHDEFEVGRDV 1740
Db 1681 TEETHVVMKTDABFCERTLKVFLGAGGWVYSYFWWTQSIKERKMLNEHDEFEVGRDV 1740
QY 1741 VNGRNHQGPKRARESDQRKIFRGLIEICCYGPFNTNMPDQLEWMVQLCGASVYKELSSFTL 1800
Db 1741 VNGRNHQGPKRARESDQRKIFRGLIEICCYGPFNTNMPDQLEWMVQLCGASVYKELSSFTL 1800
QY 1801 GTGVHPITVVVQPDWATEDNGFHAIGQCEAPVVTREWVLDSVALYQCELDLYLIPQIPH 1860
Db 1801 GTGVHPITVVVQPDWATEDNGFHAIGQCEAPVVTREWVLDSVALYQCELDLYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 7
US-08-483-554B-2
: Sequence 2, Application US/08483554B
: Patent No. 5747282
: GENERAL INFORMATION:
: APPLICANT: Skolnick, Mark H.
: APPLICANT: Goldgar, David E.
: APPLICANT: Miki, Yoshio
: APPLICANT: Swenson, Jeff
: APPLICANT: Kamb, Alexander
: APPLICANT: Harshman, Keith D.
: APPLICANT: Shattuck-Eidens, Donna M.
: APPLICANT: Tavtigian, Sean V.
: APPLICANT: Wiseman, Roger W.
: APPLICANT: Futreal, P. Andrew
: TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
: TITLE OF INVENTION: Susceptibility Gene
: NUMBER OF SEQUENCES: 85
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
: STREET: 1201 New York Avenue, N.W., Suite 1000
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/483,554B
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/409,305
: FILING DATE: 24-MAR-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/348,824
: FILING DATE: 29-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/308,104
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/300,266
: FILING DATE: 02-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/289,221
: FILING DATE: 12-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Ihnen, Jeffrey L.
: REGISTRATION NUMBER: 28,957
: REFERENCE/DOCKET NUMBER: 24884-109347
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-962-4810
: TELEFAX: 202-962-8300
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1863 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-483-554B-2

Query Match 99.9%; Score 9642; DB 1; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQNVINAMQKILECPICLETIKPEVPSTKCDHFCKFCMLKLLNOKKGPQ 60
Db 1 MDLSALRVEEVQNVINAMQKILECPICLETIKPEVPSTKCDHFCKFCMLKLLNOKKGPQ 60
QY 61 CPLCKNDITKRSIQESTRFSQLVVEELLKIICAFOLDGTGLEVANSYNFAKKENNSPEHLKD 120

Db 61 CPLECKNDITKRSLOQSTRESQLVEELKIIICAFQLDTGLEVANSYNFAKKENNSPEHLKD 120
QY 121 EVSIIOSMGVYRNRAKRLLOSEPNPSLOQTSLSVQLSNLGTVRTLRTKRIQOPKTSVYI 180
Db 121 EVSIIOSMGVYRNRAKRLLOSEPNPSLOQTSLSVQLSNLGTVRTLRTKRIQOPKTSVYI 180
QY 181 ELGSDSSEDTVNKATYCSVGDOELLOITPQGTTRDEISLSDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDTVNKATYCSVGDOELLOITPQGTTRDEISLSDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPKBYOGSSVSNLHVPCGTNTNTHASSLQENSHSLLTAKORMNVE 300
Db 241 PSNNDLNTTEKRAAERHPKBYOGSSVSNLHVPCGTNTNTHASSLQENSHSLLTAKORMNVE 300
QY 301 KAFECNKSQOPGLARSOHRNWRAGSKETCNDNRTPTSEKKYVDLNADPLCERKNWQKLPK 360
Db 301 KAFECNKSQOPGLARSOHRNWRAGSKETCNDNRTPTSEKKYVDLNADPLCERKNWQKLPK 360
QY 361 SENPRDTEVPWITLNSSTIQKYNWFSRDELIGSDSDHGESESNKAVADVLDVINEVD 420
Db 361 SENPRDTEVPWITLNSSTIQKYNWFSRDELIGSDSDHGESESNKAVADVLDVINEVD 420
QY 421 EYSGSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKXKASLPNLSHVTEN 480
Db 421 EYSGSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKXKASLPNLSHVTEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPEMINQGTNOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPEMINQGTNOTE 540
QY 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELEINI 600
Db 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELEINI 600
QY 601 HNSKAPKKNRLRRKSTRIHAELELVVSNLSPNCTELQIDSCSSSEIEIKKKYQNMVP 660
Db 601 HNSKAPKKNRLRRKSTRIHAELELVVSNLSPNCTELQIDSCSSSEIEIKKKYQNMVP 660
QY 661 RHNRNLQMEGEPATGAKSNKPNQOTSKRHDSDTFPELKLITNAPGFTKCSNTSELKE 720
Db 661 RHNRNLQMEGEPATGAKSNKPNQOTSKRHDSDTFPELKLITNAPGFTKCSNTSELKE 720
QY 721 FVNPSLPREEKEKLETKVVSNAEDPKDMLSGERVLTQTERSVESSISLVPGTDYGTQ 780
Db 721 FVNPSLPREEKEKLETKVVSNAEDPKDMLSGERVLTQTERSVESSISLVPGTDYGTQ 780
QY 781 ESISLLEVSTLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDETEGFKYPLGHEVNH 840
Db 781 ESISLLEVSTLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDETEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLONTFKVSRQSFALPSNPGNAEEECATSAHSGSLKKOSPKVT 900
Db 841 RETSIEMESELDAQYLONTFKVSRQSFALPSNPGNAEEECATSAHSGSLKKOSPKVT 900
QY 901 FECEQKEENQGNESNIKPVQTVNITAGFPVVGQKDKPDVNAKCSIKGSRFCLSSQFRG 960
Db 901 FECEQKEENQGNESNIKPVQTVNITAGFPVVGQKDKPDVNAKCSIKGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLFIKGFVKTKCKNLLLENFEBHSMSPREMGNEINP 1020
Db 961 NETGLITPNKHGLLQNPYRIPPLFIKGFVKTKCKNLLLENFEBHSMSPREMGNEINP 1020
QY 1021 STVSTISRNRIENVFKEASSNINEVGSSTNEVGSSINEIGSSDENIQOAEIAGNRGPKL 1080
Db 1021 STVSTISRNRIENVFKEASSNINEVGSSTNEVGSSINEIGSSDENIQOAEIAGNRGPKL 1080
QY 1081 NAMRLGVLOPEVYKQSLPGSNCKHPEIKKQYEEVQTVNTDFSPYILISNLEQPMGSS 1140
Db 1081 NAMRLGVLOPEVYKQSLPGSNCKHPEIKKQYEEVQTVNTDFSPYILISNLEQPMGSS 1140
QY 1141 HASQVCSETPDLDLDDGEIKEDTSAENDIKESSAVFSKSVQKGLSRSPSPFTHTHLAQ 1200

Db 1141 HASQVCSETPDLDLDDGEIKEDTSAENDIKESSAVFSKSVQKGLSRSPSPFTHTHLAQ 1200
QY 1201 GYRRGAKKLESSEENLSSEDEELPCFOHLLFGKVNIPSQSTRHSHTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSSEDEELPCFOHLLFGKVNIPSQSTRHSHTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDSCNOVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLNDSCNOVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHQSESQGVGLSKDELVSDDERGTGLEENNOEOMSDSNLGEAASGESETSVSE 1380
Db 1321 SKQMRHQSESQGVGLSKDELVSDDERGTGLEENNOEOMSDSNLGEAASGESETSVSE 1380
QY 1381 DCSGLSSOSDILTTQORDTMQHNLIKLOQEMAELEAVLEOHGSQPSNSYPSIISDSALE 1440
Db 1381 DCSGLSSOSDILTTQORDTMQHNLIKLOQEMAELEAVLEOHGSQPSNSYPSIISDSALE 1440
QY 1441 DLNRNPEQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSSSTSKNKEPGVERSPSPK 1500
Db 1441 DLNRNPEQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSSSTSKNKEPGVERSPSPK 1500
QY 1501 CPSLDDRWMHSCSGSLQNRNYPQSEELIKVVDVEEQQLLESGPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDDRWMHSCSGSLQNRNYPQSEELIKVVDVEEQQLLESGPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESGISLSDPESDPEDRAPESARVGNIPSTTSALKVPOLKVAESAQSPAAHHT 1620
Db 1561 TPYLESGISLSDPESDPEDRAPESARVGNIPSTTSALKVPOLKVAESAQSPAAHHT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSMVVSGLTPPEEFMLVYKFARKHHITLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSMVVSGLTPPEEFMLVYKFARKHHITLTNLI 1680
QY 1681 TEETHVVMYKTDAEFVCERTIKYFLGIAGGKVVVYFWVTQSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMYKTDAEFVCERTIKYFLGIAGGKVVVYFWVTQSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQPKRARESODRKIPRGLEICCYGPTNPTDOLEHNVOLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQPKRARESODRKIPRGLEICCYGPTNPTDOLEHNVOLCGASVVKELSSFTL 1800
QY 1801 GTGVHPIVVVPDAMTNGFHAIGQMCEAPVVTREWVLDLSVALYQCQELDYLIPIQIPH 1860
Db 1801 GTGVHPIVVVPDAMTNGFHAIGQMCEAPVVTREWVLDLSVALYQCQELDYLIPIQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 8

US-08-488-011B-2
; Sequence 2, Application US/08488011B
; Patent No. 5753441
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Mikil, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington

STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,011B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Innen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347-09
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-488-011B-2

Query Match 99.9%; Score 9642; DB 1; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	421	EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTEN	480
DB	421	EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTEN	480
QY	481	LIIGAFVTEPQIIQERPLTNKLKRKRPTSGLHPDEFIKKADLAVQKTPEMINOQNTQTE	540
DB	481	LIIGAFVTEPQIIQERPLTNKLKRKRPTSGLHPDEFIKKADLAVQKTPEMINOQNTQTE	540
QY	541	ONGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI	600
DB	541	ONGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI	600
QY	601	HNSKAPKKNRLRRKSSTRIHALELVVSRNLSPNCTELQIDSCSSSEIKKKKYNQMPV	660
DB	601	HNSKAPKKNRLRRKSSTRIHALELVVSRNLSPNCTELQIDSCSSSEIKKKKYNQMPV	660
QY	661	RHSRNLQMEGKEPATGAKKSNKPNEQTSKRHDSDFPELKLITNAPGSKTCSNTSELKE	720
DB	661	RHSRNLQMEGKEPATGAKKSNKPNEQTSKRHDSDFPELKLITNAPGSKTCSNTSELKE	720
QY	721	FYNPSLPREEKEELKLVKVSNNADPKDMLSGERVLOTERSVSSSISLVPDGYGTQ	780
DB	721	FYNPSLPREEKEELKLVKVSNNADPKDMLSGERVLOTERSVSSSISLVPDGYGTQ	780
QY	781	ESISLLEVTGLGKAKTEPNKCVSQAAPENPKGLIHGCCKDNRNDEGFKYPLGHEVNH	840
DB	781	ESISLLEVTGLGKAKTEPNKCVSQAAPENPKGLIHGCCKDNRNDEGFKYPLGHEVNH	840
QY	841	RETSIEMESELDAQYLQNTFKVSKRQSPALFSPNGNAEEECATPSAHSGLSKKQSPKVT	900
DB	841	RETSIEMESELDAQYLQNTFKVSKRQSPALFSPNGNAEEECATPSAHSGLSKKQSPKVT	900
QY	901	FECEQKEENQGNESNIKPVQVNVITAGFPVVGQKDKPVDNAKCSIKGSRCLSSQFRG	960
DB	901	FECEQKEENQGNESNIKPVQVNVITAGFPVVGQKDKPVDNAKCSIKGSRCLSSQFRG	960
QY	961	NETGLITPNKHGLLQNPYRIPPLFKPSFKVTCKKNLLEENFEHSHMSPEREMGNENTP	1020
DB	961	NETGLITPNKHGLLQNPYRIPPLFKPSFKVTCKKNLLEENFEHSHMSPEREMGNENTP	1020
QY	1021	STVSTISRNNIRENVFKEASSNINEVSGSTNEVGSSSINEIGSSDENIQAELGRNRPKL	1080
DB	1021	STVSTISRNNIRENVFKEASSNINEVSGSTNEVGSSSINEIGSSDENIQAELGRNRPKL	1080
QY	1081	NAMRLGVLOPEVYKOSLPGSNCKHPKIKQYEEVQVNTDFSPYLISDNLEQPMGSS	1140
DB	1081	NAMRLGVLOPEVYKOSLPGSNCKHPKIKQYEEVQVNTDFSPYLISDNLEQPMGSS	1140
QY	1141	HASQVCSETPDDLLDDGEIKEDTSAENDIKESSAVFSKSVOKGELSRSPSPFTHHLAQ	1200
DB	1141	HASQVCSETPDDLLDDGEIKEDTSAENDIKESSAVFSKSVOKGELSRSPSPFTHHLAQ	1200
QY	1201	GYRGAKKLESSEENLSSEDEELPCFQHLHFGKVNINIPSQSTRHSTVATECLSKNTEENL	1260
DB	1201	GYRGAKKLESSEENLSSEDEELPCFQHLHFGKVNINIPSQSTRHSTVATECLSKNTEENL	1260
QY	1261	LSLKNLSDCNSQVITLAKASQEHHLSEETKCSASFSSOCSELEDLTANTNTQDPFLIGS	1320
DB	1261	LSLKNLSDCNSQVITLAKASQEHHLSEETKCSASFSSOCSELEDLTANTNTQDPFLIGS	1320
QY	1321	SKQMRHOSQGVGSLDKELVSDDERGTGLEENQOEQSDNSNLGEAAGCESSETSVSE	1380
DB	1321	SKQMRHOSQGVGSLDKELVSDDERGTGLEENQOEQSDNSNLGEAAGCESSETSVSE	1380
QY	1381	DCSGLSSQSDILTTQORDTMQHNLKIQOEMAELEAVLEQHGHSQSPNSYPSIISDSSALE	1440
DB	1381	DCSGLSSQSDILTTQORDTMQHNLKIQOEMAELEAVLEQHGHSQSPNSYPSIISDSSALE	1440
QY	1441	DLRNPQSTSEKAVLTQSQSSSEYPIQSNPEGLSADKFEVSADSSSTSKNKEPGVERSSPSK	1500
DB	1441	DLRNPQSTSEKAVLTQSQSSSEYPIQSNPEGLSADKFEVSADSSSTSKNKEPGVERSSPSK	1500

```
QY 1501 CPSLDRWYMHSCSGSLQNRNYPQSEELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDRWYMHSCSGSLQNRNYPQSEELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESGISLFDSPESDRAPEARSARVGNIPSTTSALKVPQKVAESAQSPAAAHHT 1620
Db 1561 TPYLESGISLFDSPESDRAPEARSARVGNIPSTTSALKVPQKVAESAQSPAAAHHT 1620
QY 1621 DTAGYNAMEESYSRKPELTASTERVNKRMSMVVSGLTPEEFMLVYKPAKHHITLTNLI 1680
Db 1621 DTAGYNAMEESYSRKPELTASTERVNKRMSMVVSGLTPEEFMLVYKPAKHHITLTNLI 1680
QY 1681 TEETHVVMKTDAEVCERTLKYFLGIAGGKVVSYFWVTQSIKIRKMLNHDHFEVRGDV 1740
Db 1681 TEETHVVMKTDAEVCERTLKYFLGIAGGKVVSYFWVTQSIKIRKMLNHDHFEVRGDV 1740
QY 1741 VNGRNHQGPKRAESQDRKIFRGLBICCYGPTNNPTDQLEWMVQLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQGPKRAESQDRKIFRGLBICCYGPTNNPTDQLEWMVQLCGASVVKELSSFTL 1800
QY 1801 GTGVHPVVVQPDAMTEDNGFHAIGOMCEAPVVTREWVLDSVALYQCOELDTYLIPOLPH 1860
Db 1801 GTGVHPVVVQPDAMTEDNGFHAIGOMCEAPVVTREWVLDSVALYQCOELDTYLIPOLPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 9
US-08-850-727-2
; Sequence 2, Application US/08850727
; Patent No. 6162897
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/850,727
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/483,554
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-850-727-2

Query Match 99.98; Score 9642; DB 4; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSARVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNOKKGPSQ 60
Db 1 MDLSARVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNOKKGPSQ 60
QY 61 CPLCKNDITKRSLOESTRFSQVBEELKIIICAFOLDTGLEVANSYNFAKKNNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVBEELKIIICAFOLDTGLEVANSYNFAKKNNSPEHLKD 120
QY 121 EVSIIQSMGYNRNRAKRLLOSPEPENPSLQETSLSVQLSNLGTVRTLTQRTQPKQTSYVI 180
Db 121 EVSIIQSMGYNRNRAKRLLOSPEPENPSLQETSLSVQLSNLGTVRTLTQRTQPKQTSYVI 180
QY 181 ELGSDSSEDVTNKATYCSVGDOELLQITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDVTNKATYCSVGDOELLQITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPEKYQGVSSVNLHVPCGTNTHASSLOHENSLLLTQDKRMNVE 300
Db 241 PSNNDLNTTEKRAAERHPEKYQGVSSVNLHVPCGTNTHASSLOHENSLLLTQDKRMNVE 300
QY 301 KAEFCNKSQKQGLARSOHNRWAGSKETCNDRTPTSTKKVDLNDADPLCERKEWNNKQLPC 360
Db 301 KAEFCNKSQKQGLARSOHNRWAGSKETCNDRTPTSTKKVDLNDADPLCERKEWNNKQLPC 360
QY 361 SENPRDTEVPWITLNSSIQVNEWFSRSDLLGSDSDSHDGESESNKAVADVLDVINEVD 420
Db 361 SENPRDTEVPWITLNSSIQVNEWFSRSDLLGSDSDSHDGESESNKAVADVLDVINEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKYRKASLPNLSHVTEEN 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKYRKASLPNLSHVTEEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFTKKADLAVQKTPMINOCTNOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFTKKADLAVQKTPMINOCTNOTE 540
QY 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTLESLEKESAFKTKAEPISSSISNMELELNI 600
Db 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTLESLEKESAFKTKAEPISSSISNMELELNI 600
QY 601 HNSKAPKKNLRLRKSSRTHIALELVYSRNLSPNCTELQIDSCSSSBEIKKKYQNPV 660
Db 601 HNSKAPKKNLRLRKSSRTHIALELVYSRNLSPNCTELQIDSCSSSBEIKKKYQNPV 660
QY 661 RHRNRLQMEGKEPATGAKKSNKNEQTSKRHSDTPELKLTNAPGSFTKCSNTSELKE 720
Db 661 RHRNRLQMEGKEPATGAKKSNKNEQTSKRHSDTPELKLTNAPGSFTKCSNTSELKE 720
QY 721 FVNPSPREEKEKLETYKVSNNADPKDMLSGSERVLTQTERSVESSSISLVPQTDYGTQ 780
```

c

Db 721 FVNPSPREEKEKLETVKYSVNAEDPKDMLSGERVLOTERSVESSTSLVPGTDYGTQ 780
QY 781 ESISLLEVSTLGKAKTEPNKCSQCAAFENPKGLIHGCCSDNRNDTEGPKYPLGHEVNH 840
Db 781 ESISLLEVSTLGKAKTEPNKCSQCAAFENPKGLIHGCCSDNRNDTEGPKYPLGHEVNH 840
QY 841 RETSIEMEESLDAOYLQNTFKYVKRQSFALFSPNCAEBECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMEESLDAOYLQNTFKYVKRQSFALFSPNCAEBECATFSAHSGSLKKQSPKVT 900
QY 901 FECEQKEENQGNESIKPVQTVNITAGPVVQKDKPDVNAKCSIKGGSRFLCSQFRG 960
Db 901 FECEQKEENQGNESIKPVQTVNITAGPVVQKDKPDVNAKCSIKGGSRFLCSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLFPKTSVTKCKNLLLEENPEEHSMPEREMGNENIP 1020
Db 961 NETGLITPNKHGLLQNPYRIPPLFPKTSVTKCKNLLLEENPEEHSMPEREMGNENIP 1020
QY 1021 STVSTISRNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNRPKL 1080
Db 1021 STVSTISRNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNRPKL 1080
QY 1081 NAMLRGLVLOPEVYKOSLPGSNCKHPEIKKQYEEVVQTVNTDFSPYLSIDNLEQPMGSS 1140
Db 1081 NAMLRGLVLOPEVYKOSLPGSNCKHPEIKKQYEEVVQTVNTDFSPYLSIDNLEQPMGSS 1140
QY 1141 HASOVCSPTDLDLDDGETIKEDTSFAENDIKESSAVFSKVGKELSRSPSPHTHQAQ 1200
Db 1141 HASOVCSPTDLDLDDGETIKEDTSFAENDIKESSAVFSKVGKELSRSPSPHTHQAQ 1200
QY 1201 GYRRGAKKLESSEENLSSEDEELPCQHLFGKVNIPQSQRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSSEDEELPCQHLFGKVNIPQSQRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLDCSNQVILAKASQEHLSSEETKCSASLFSQCSSELEDJTANTNTQDPFLIG 1320
Db 1261 LSLKNSLDCSNQVILAKASQEHLSSEETKCSASLFSQCSSELEDJTANTNTQDPFLIG 1320
QY 1321 SKOMRHQESQGVGLSDKELVDDEERGTLGLENNOEQSMDSNLGEAASGESETSVSE 1380
Db 1321 SKOMRHQESQGVGLSDKELVDDEERGTLGLENNOEQSMDSNLGEAASGESETSVSE 1380
QY 1381 DCSGLSSQSDILTTQORDTMQHNLIKLOQEMAELEAVLEHQSGSQSPNSYPSIISDSSALE 1440
Db 1381 DCSGLSSQSDILTTQORDTMQHNLIKLOQEMAELEAVLEHQSGSQSPNSYPSIISDSSALE 1440
QY 1441 DLNRPQSTSEKAVLTQSQSSSEYPISONPEGLSADKFEVSADSSTSKNKEPGVERSSPSK 1500
Db 1441 DLNRPQSTSEKAVLTQSQSSSEYPISONPEGLSADKFEVSADSSTSKNKEPGVERSSPSK 1500
QY 1501 CPSLDRWTFMHSQSGSLQNRNYPQBELTKVVDVVEQQLLEESGPHDLTTSYLPQDLEG 1560
Db 1501 CPSLDRWTFMHSQSGSLQNRNYPQBELTKVVDVVEQQLLEESGPHDLTTSYLPQDLEG 1560
QY 1561 TPYLESGISLFDSDPDSPEDRAPESARVGNIPSSTSALKVPQLKVAESAQSPAAAHPT 1620
Db 1561 TPYLESGISLFDSDPDSPEDRAPESARVGNIPSSTSALKVPQLKVAESAQSPAAAHPT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEPMVLYKFARKKHITLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEPMVLYKFARKKHITLTNLI 1680
QY 1681 TETTHVMKTDAEFCERTLKYFLGIAGGKVVVSVFWTQSTIKERKMLNEHDFEVRGDV 1740
Db 1681 TETTHVMKTDAEFCERTLKYFLGIAGGKVVVSVFWTQSTIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQPKRARESODRKIFRGLEICCCGPTNPTDQLEWVVLCSGVVKELSSFTL 1800
Db 1741 VNGRNHQPKRARESODRKIFRGLEICCCGPTNPTDQLEWVVLCSGVVKELSSFTL 1800
QY 1801 GTGVHPVVVVQPDAWTNGFHAIGQMCAPVVTREWLDSVALYQCQELDTYLIPQIPH 1860

Db 1801 GTGVHPVVVVQPDAWTNGFHAIGQMCAPVVTREWLDSVALYQCQELDTYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863
RESULT 10
PCT-US95-10202-2
; Sequence 2, Application PC/TUS9510202
; GENERAL INFORMATION:
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Simard, Jacques
; APPLICANT: Emi, Mitsuru
; APPLICANT: Nakamura, Yusuke
; APPLICANT: Durocher, Francine
; TITLE OF INVENTION: In Vivo Mutations and Polymorphisms
; TITLE OF INVENTION: in the 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10202
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08-308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-10202-2
Query Match 99.9%; Score 9642; DB 5; Length 1863;
Best Local Similarity 99.9%; Pred. NO. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQVNIAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKGPSQ 60
Db 1 MDLSALRVEEVQVNIAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKGPSQ 60
QY 61 CPLCKNDITKKSLOBSTFSQLVEELLKICAFQLDGTGLEAYANSYNPAKKENNSPEHLKD 120
Db 61 CPLCKNDITKKSLOBSTFSQLVEELLKICAFQLDGTGLEAYANSYNPAKKENNSPEHLKD 120
QY 121 EYSIIQSOGYRNRAKRLQSPENPSLQETSLSVOLSNLGNVRLTRTKORTQPOKTSVYI 180
Db 121 EYSIIQSOGYRNRAKRLQSPENPSLQETSLSVOLSNLGNVRLTRTKORTQPOKTSVYI 180
QY 181 ELGSDSSEDVTNKATYCSVGQOELLQITPQGTREDEISLDSAKKAACBFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDVTNKATYCSVGQOELLQITPQGTREDEISLDSAKKAACBFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTERAAARRHPKTYQGSVSVMHLVPECGTNTHASSLQHENSSLLLTTKDRMNVE 300
Db 241 PSNNDLNTTERAAARRHPKTYQGSVSVMHLVPECGTNTHASSLQHENSSLLLTTKDRMNVE 300
QY 301 KAEFCNKSQPCGLARSOHNRWAGSKETCNDRRTPSTKKYVDLNDADPLCERKEWNKQKLP 360
Db 301 KAEFCNKSQPCGLARSOHNRWAGSKETCNDRRTPSTKKYVDLNDADPLCERKEWNKQKLP 360
QY 361 SENPRDTEVPWITLNSIQVNEWFSRSDLLGSDSDSHDGESESNAKVADVLDVLNEVD 420
Db 361 SENPRDTEVPWITLNSIQVNEWFSRSDLLGSDSDSHDGESESNAKVADVLDVLNEVD 420
QY 421 EYSGSSEKIDLLASDPHALLCKSERVHKSVSNEIEDKIFGKYRKKASLPNLSHVYEN 480
Db 421 EYSGSSEKIDLLASDPHALLCKSERVHKSVSNEIEDKIFGKYRKKASLPNLSHVYEN 480
QY 481 LIIGAFVTEPOLIOBRPLTNLKKRRTPSGLHPEDFTKKRADLAVOKTPEMINOGTNOTE 540
Db 481 LIIGAFVTEPOLIOBRPLTNLKKRRTPSGLHPEDFTKKRADLAVOKTPEMINOGTNOTE 540
QY 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTESLEKESAFKTKAEPITSSISNMELELNI 600
Db 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTESLEKESAFKTKAEPITSSISNMELELNI 600
QY 601 HNSKAPKNRLRRKSSSTRIHAILVELVSRNLSPPNCTELQIDSCSSSEEEIKKKYINOMPV 660
Db 601 HNSKAPKNRLRRKSSSTRIHAILVELVSRNLSPPNCTELQIDSCSSSEEEIKKKYINOMPV 660
QY 661 RHSRNLQIMECKEPATGAKKSNKNEOTSKRHDSDTPELKLTNAPGSFTKCSNTSELKE 720
Db 661 RHSRNLQIMECKEPATGAKKSNKNEOTSKRHDSDTPELKLTNAPGSFTKCSNTSELKE 720
QY 721 FVNPSLPREEKEEKLETYKVSNNNAEDPKDMLSGERVLOTERSVESSISILVPGTDYGTQ 780
Db 721 FVNPSLPREEKEEKLETYKVSNNNAEDPKDMLSGERVLOTERSVESSISILVPGTDYGTQ 780
QY 781 ESIISLLEVSTLIGKAKTENPKVSOCAAFENPKGLIHGCSKDNDRNDTEGFKYPLGHEVNH 840
Db 781 ESIISLLEVSTLIGKAKTENPKVSOCAAFENPKGLIHGCSKDNDRNDTEGFKYPLGHEVNH 840
QY 841 RETSTEMESELDAQYLQNTFVKSKROSFAFVSNPGNAEEECATFSAHSGSLKKOSPKVT 900
Db 841 RETSTEMESELDAQYLQNTFVKSKROSFAFVSNPGNAEEECATFSAHSGSLKKOSPKVT 900
QY 901 FECEQKEENQCKNESNTKPVQTVNITAGFPVVGQKDKPVDNAKCSIKGGSRFCILSSQFRG 960
Db 901 FECEQKEENQCKNESNTKPVQTVNITAGFPVVGQKDKPVDNAKCSIKGGSRFCILSSQFRG 960
QY 961 NETGLITPNKKHGLLQNPYRIPPLFPKSFVKTKCKKNLLEENFEHSMSPEREMGNENIP 1020
Db 961 NETGLITPNKKHGLLQNPYRIPPLFPKSFVKTKCKKNLLEENFEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNNTRENVPFKAASSNINEVGSSTNEVSGSSINEIGSDDENIQAEIAGLRNRPGL 1080
Db 1021 STVSTISRNNTRENVPFKAASSNINEVGSSTNEVSGSSINEIGSDDENIQAEIAGLRNRPGL 1080
QY 1081 NAMLRGLVQPEVYKQSLPGSNCKHPEIKKQYEVEVQTVNTDFSPYLIISDNLEQPMGSS 1140

Db 1081 NAMLRGLVQPEVYKQSLPGSNCKHPEIKKQYEVEVQTVNTDFSPYLIISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDDLLDGDGEIKEDETSFAENDIKESSAVFSKSVOKGELSRSPPFTHTHLAQ 1200
Db 1141 HASQVCSETPDDLLDGDGEIKEDETSFAENDIKESSAVFSKSVOKGELSRSPPFTHTHLAQ 1200
QY 1201 GYRRGAKKLSESEENLSSEDEELPCFQHLFLGKVNNTPSQSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLSESEENLSSEDEELPCFQHLFLGKVNNTPSQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHOSQSGVGLSDKELVSDDEERGTGLERNQBEQSDMNLGEAASCESETSYSE 1380
Db 1321 SKQMRHOSQSGVGLSDKELVSDDEERGTGLERNQBEQSDMNLGEAASCESETSYSE 1380
QY 1381 DCSGLSSQSDILTTQORDTMQHNLIKLOQEMAELEAVLEQHGQSPSNSYPISIISDSSALE 1440
Db 1381 DCSGLSSQSDILTTQORDTMQHNLIKLOQEMAELEAVLEQHGQSPSNSYPISIISDSSALE 1440
QY 1441 DLRNPEQSTSEKAVLTSSQKSEYPISONPEGLSADKFEVSADSSSTSKNKEPVERSSPSK 1500
Db 1441 DLRNPEQSTSEKAVLTSSQKSEYPISONPEGLSADKFEVSADSSSTSKNKEPVERSSPSK 1500
QY 1501 CPSLDDRWYMHSCSGSLQNRNYPISQBELIKVVDVEEQOLESGBPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDDRWYMHSCSGSLQNRNYPISQBELIKVVDVEEQOLESGBPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESIGISLFSDDPESDRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT 1620
Db 1561 TPYLESIGISLFSDDPESDRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVYSGITPTEEFMLVYKPAKHHHTLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVYSGITPTEEFMLVYKPAKHHHTLTNLI 1680
QY 1681 TEETHVVMKTDAEFVCERTLKYPFLGTAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDAEFVCERTLKYPFLGTAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQGPARRAESQDRKIFRGLIEICCYGPFNTMPDQLEWMVQLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQGPARRAESQDRKIFRGLIEICCYGPFNTMPDQLEWMVQLCGASVVKELSSFTL 1800
QY 1801 GTGVHPITVVVQPDWNTEDNGFHAIGQMCAPVVTREWVLDVSVALYQCOELDTYLIQIPH 1860
Db 1801 GTGVHPITVVVQPDWNTEDNGFHAIGQMCAPVVTREWVLDVSVALYQCOELDTYLIQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863
RESULT 11
PCT-US95-10203-2
; Sequence 2, Application PC/TUS9510203
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Bidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10203
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08-308,104
FILING DATE: 16-SEP-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Innen, Jeffrey L.
REGISTRATION NUMBER: 28,957

REFERENCE/DOCKET NUMBER: 24884-109347

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-962-4810

TELEFAX: 202-962-8300

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1863 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US95-10203-2

Query Match 99.9%; Score 9642; DB 5; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MDLSALRVEEVQVNVINAMQKILECPICLLEIKPEVPSTKCDHIFCKFCMLKLLNQKKGPSQ	60
DB	1	MDLSALRVEEVQVNVINAMQKILECPICLLEIKPEVPSTKCDHIFCKFCMLKLLNQKKGPSQ	60
QY	61	CPICKNDITKRSLOESTRFSQVVEELLKIIICAFQDITGLYANSYNFACKENNSPEHLKD	120
DB	61	CPICKNDITKRSLOESTRFSQVVEELLKIIICAFQDITGLYANSYNFACKENNSPEHLKD	120
QY	121	EVSIIQSMGYRNRAKLLQSEPNPSLQETSLSVQLSNLGTVRTLTORKIOPQKTSVYI	180
DB	121	EVSIIQSMGYRNRAKLLQSEPNPSLQETSLSVQLSNLGTVRTLTORKIOPQKTSVYI	180
QY	181	ELGSDSDSDTVNKATYCSVDQELLQITPQGTREISLDSAKKAACEFSETDVTNTEHHQ	240
DB	181	ELGSDSDSDTVNKATYCSVDQELLQITPQGTREISLDSAKKAACEFSETDVTNTEHHQ	240
QY	241	PSNNDLNTTEKRAAERHPKPYGSSVSNLHVPCGNTNTHASSLQHENSLLLTKDRMNYE	300
DB	241	PSNNDLNTTEKRAAERHPKPYGSSVSNLHVPCGNTNTHASSLQHENSLLLTKDRMNYE	300

QY	301	KAFCNCKSQPGLARSQHNWAGSKETCNDRRTPSTPEKKVLDLNADPLCERKKNQKLP	360
DB	301	KAFCNCKSQPGLARSQHNWAGSKETCNDRRTPSTPEKKVLDLNADPLCERKKNQKLP	360
QY	361	SENPRDTEVPWITLNSSIQKVNEWFSRDELGSDSDSHDGESENAKVADVLVDLNEVD	420
DB	361	SENPRDTEVPWITLNSSIQKVNEWFSRDELGSDSDSHDGESENAKVADVLVDLNEVD	420
QY	421	EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN	480
DB	421	EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN	480
QY	481	LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPDEFIKKADLAVQKTPEMINQGTNQE	540
DB	481	LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPDEFIKKADLAVQKTPEMINQGTNQE	540
QY	541	QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKKAEPISISSINMELELNI	600
DB	541	QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKKAEPISISSINMELELNI	600
QY	601	HNSKAPKKNRLRRKSSSTRHIALELVVSRLNPPNCTELQIDSCSSSEETKKKKYNQMPV	660
DB	601	HNSKAPKKNRLRRKSSSTRHIALELVVSRLNPPNCTELQIDSCSSSEETKKKKYNQMPV	660
QY	661	RHSRLQLMEGKEPATGAKKSNKPNQTSKRHSDSTFPELKLNTNAPGSKTCSNTSELKE	720
DB	661	RHSRLQLMEGKEPATGAKKSNKPNQTSKRHSDSTFPELKLNTNAPGSKTCSNTSELKE	720
QY	721	FVNPSLPREEKEKLETVKVSNNADDPKDLMLSGERVLQTERSVSSSSISLVPGTDYGTQ	780
DB	721	FVNPSLPREEKEKLETVKVSNNADDPKDLMLSGERVLQTERSVSSSSISLVPGTDYGTQ	780
QY	781	ESISLLEVTGLKAKTEPNKCVSOCAAFENPKGLIHGCCSDNRNDTEGKYPLGHEVNIH	840
DB	781	ESISLLEVTGLKAKTEPNKCVSOCAAFENPKGLIHGCCSDNRNDTEGKYPLGHEVNIH	840
QY	841	RETSIEMEESSELDQYLQNTFKVSRQSPALFSPNCPNABEECATFSAHSGSLKKQSPKVT	900
DB	841	RETSIEMEESSELDQYLQNTFKVSRQSPALFSPNCPNABEECATFSAHSGSLKKQSPKVT	900
QY	901	FECEQKEENQGNESNIKPVQTVNITAGPPVVGQDKDPVDNAKCSIKGSRCLSSQFRG	960
DB	901	FECEQKEENQGNESNIKPVQTVNITAGPPVVGQDKDPVDNAKCSIKGSRCLSSQFRG	960
QY	961	NETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKNLLLEENFEHSMSPERMGNEINIP	1020
DB	961	NETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKNLLLEENFEHSMSPERMGNEINIP	1020
QY	1021	STVSTISRNNIRENVFKEASSNINEVGSSTNEVGSSSINEIGSSDENIOAELGRNRPKL	1080
DB	1021	STVSTISRNNIRENVFKEASSNINEVGSSTNEVGSSSINEIGSSDENIOAELGRNRPKL	1080
QY	1081	NAMLRLGVLOPEVYKOSLPGSNCKHPEIKKQYEEVVQVTVNTDFSPYLSLDNLEQPMGSS	1140
DB	1081	NAMLRLGVLOPEVYKOSLPGSNCKHPEIKKQYEEVVQVTVNTDFSPYLSLDNLEQPMGSS	1140
QY	1141	HASQVCSETPDLLDDGEIKEDTSAENDIKESSAVFSKVQKGLSRSPSFTHHLAQ	1200
DB	1141	HASQVCSETPDLLDDGEIKEDTSAENDIKESSAVFSKVQKGLSRSPSFTHHLAQ	1200
QY	1201	GYYRGAKKLSESEENLSSEDEELPCFOHLLFGKVNNIPSQSTRHSTVATECLSKNTEENL	1260
DB	1201	GYYRGAKKLSESEENLSSEDEELPCFOHLLFGKVNNIPSQSTRHSTVATECLSKNTEENL	1260
QY	1261	LSLKNSLNDCSNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS	1320
DB	1261	LSLKNSLNDCSNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS	1320
QY	1321	SKQMRHQSESQGVGLSDKELVSDDEERGTCLEENQEQESMDNSNLGEAASGESETSVSE	1380
DB	1321	SKQMRHQSESQGVGLSDKELVSDDEERGTCLEENQEQESMDNSNLGEAASGESETSVSE	1380
QY	1381	DCSGLSSQSDILTTQOQRTMQHNLKIQEQEMAELEAVLHQGSQPSNSYPSIISDSSALE	1440

1381 DCSGLSSQSDILTTQORDTMOHNLKLOQMAEAVLEQHQSPNSYPSIISDSSALE 1440
1441 DURNPQSTSEKAVLTQKSSSEYPIQSONPEGLSADKFEVSADSTSKNKEPGVERSSPSK 1500
1441 DURNPQSTSEKAVLTQKSSSEYPIQSONPEGLSADKFEVSADSTSKNKEPGVERSSPSK 1500
1501 CFSLDRTWMHSCGSLQNRNYPQBELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
1501 CFSLDRTWMHSCGSLQNRNYPQBELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
1561 TPYLESIGISLSDPESDSEDRAPEARVGNIPSTGALKVPOLKVAESAQSPAAAHHT 1620
1561 TPYLESIGISLSDPESDSEDRAPEARVGNIPSTGALKVPOLKVAESAQSPAAAHHT 1620
1621 DTAGYNAMESVSRKPELTASTERVNKRMSVMVSGLTPPEFMLVYKFAKHHITLNL 1680
1621 DTAGYNAMESVSRKPELTASTERVNKRMSVMVSGLTPPEFMLVYKFAKHHITLNL 1680
1681 TEETHVVMKTDAEFCERTLKIFYLGAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
1681 TEETHVVMKTDAEFCERTLKIFYLGAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
1741 VNGRNHQPRAESQDRKIFRGLICCYGPTNNMPTDQLEMMVQLCGASVVKELSSFTL 1800
1741 VNGRNHQPRAESQDRKIFRGLICCYGPTNNMPTDQLEMMVQLCGASVVKELSSFTL 1800
1801 GTGVHPVIVVQPDWNTDNGFHAIGOMCEAPVVTREWLDSDVALYQCOELDTYLIPQIPH 1860
1801 GTGVHPVIVVQPDWNTDNGFHAIGOMCEAPVVTREWLDSDVALYQCOELDTYLIPQIPH 1860
1861 SHY 1863
1861 SHY 1863

RESULT 12

PCT-US95-10220-2
Sequence 2. Application PC/TUS9510220
GENERAL INFORMATION:
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: Method for diagnosing a
TITLE OF INVENTION: Predisposition for Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10220
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08-308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-10220-2

Query Match 99.9%; Score 9642; DB 5; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQNVINAMQKILECPICILEIKPEVSTKCDHIFCKFCKMLLNQKKGPSQ 60
DB 1 MDLSALRVEEVQNVINAMQKILECPICILEIKPEVSTKCDHIFCKFCKMLLNQKKGPSQ 60
QY 61 CPICKNDITKRSLOESTRFESOLVEELKIIICAFOLDTGLEYANSYNFAKKENNSPEHLKD 120
DB 61 CPICKNDITKRSLOESTRFESOLVEELKIIICAFOLDTGLEYANSYNFAKKENNSPEHLKD 120
QY 121 EVSIIQSMGYNRNRAKRLQSEPNPSLQETSLSVQLSNLGTVRTLTQRIQPKQTSYVI 180
DB 121 EVSIIQSMGYNRNRAKRLQSEPNPSLQETSLSVQLSNLGTVRTLTQRIQPKQTSYVI 180
QY 181 ELGSDSSSDIVNKATYCSVGDOELLQITPQGTREISLDSAKKACSEFSETDVTNTEHHQ 240
DB 181 ELGSDSSSDIVNKATYCSVGDOELLQITPQGTREISLDSAKKACSEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPEKYQGSVSNLHVPCGTNTTHASSLOHENSLLLTDRMNV 300
DB 241 PSNNDLNTTEKRAAERHPEKYQGSVSNLHVPCGTNTTHASSLOHENSLLLTDRMNV 300
QY 301 KAEFCNKSQKPLARSQHRNHWAGSKETCNDRTPTSTKVKVDNADPLCERKEWNKQKLPC 360
DB 301 KAEFCNKSQKPLARSQHRNHWAGSKETCNDRTPTSTKVKVDNADPLCERKEWNKQKLPC 360
QY 361 SENPRDTEVPWITLNSSIQKVNWFPSRDELGSDSDHDSSESNAKVADVLVDNEVD 420
DB 361 SENPRDTEVPWITLNSSIQKVNWFPSRDELGSDSDHDSSESNAKVADVLVDNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTE 480
DB 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTE 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMENOQTNQTE 540
DB 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMENOQTNQTE 540
QY 541 ONGOVMMITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPITSSSISNMELELNI 600
DB 541 ONGOVMMITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPITSSSISNMELELNI 600

QY 601 HNSKAPKNNLRKSRTHALELVVSRNLSPNCTELQIDSCSSSEEEKKKYNQMPV 660
Db 601 HNSKAPKNNLRKSRTHALELVVSRNLSPNCTELQIDSCSSSEEEKKKYNQMPV 660
QY 661 RHRNQLMEGKBPATCAKSNKPNQTSKRHSDTFPELKLTPNAGSFTKCSNTSELKE 720
Db 661 RHRNQLMEGKBPATCAKSNKPNQTSKRHSDTFPELKLTPNAGSFTKCSNTSELKE 720
QY 721 FVNPSPREEKEKLETVKYSNNAEDPKDMLSGERVLOQTERSVESSISLVPCTDYGTO 780
Db 721 FVNPSPREEKEKLETVKYSNNAEDPKDMLSGERVLOQTERSVESSISLVPCTDYGTO 780
QY 781 ESTSLLEVSTLGAKAPENPKVSCAAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
Db 781 ESTSLLEVSTLGAKAPENPKVSCAAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMEESELDQAQYLQNTFKYKQSFALFSPNPNAGAECECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMEESELDQAQYLQNTFKYKQSFALFSPNPNAGAECECATFSAHSGSLKKQSPKVT 900
QY 901 FECEQKEENOGKKNESNIPQVTNITAGFPVVGQKQKPDVNAKCSIKGSRFCLSSQFRG 960
Db 901 FECEQKEENOGKKNESNIPQVTNITAGFPVVGQKQKPDVNAKCSIKGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLPFTKSFVTKCKKNLLEENFEHSMSPEREMGNENIP 1020
Db 961 NETGLITPNKHGLLQNPYRIPPLPFTKSFVTKCKKNLLEENFEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNNIRENVKPEASSNINEVGSSTNEVGSINIEIGSSDENIQAEELGRNGPKL 1080
Db 1021 STVSTISRNNIRENVKPEASSNINEVGSSTNEVGSINIEIGSSDENIQAEELGRNGPKL 1080
QY 1081 NAMLRGLVLOPEYKQSLPGSNCKHPEIKQOEYEVVQTVNTDFSPYLLSDNLEQPMGSS 1140
Db 1081 NAMLRGLVLOPEYKQSLPGSNCKHPEIKQOEYEVVQTVNTDFSPYLLSDNLEQPMGSS 1140
QY 1141 HASQVCSPTDDLLDDGETIKEDTSAFENDIKESSAVFSKVQKGLSRSPFTHTHLAQ 1200
Db 1141 HASQVCSPTDDLLDDGETIKEDTSAFENDIKESSAVFSKVQKGLSRSPFTHTHLAQ 1200
QY 1201 GYRGAKKLESSEENLSSEDEELPCQHLFLGKVNIPQSOTRHSIVATECLSKNTEENL 1260
Db 1201 GYRGAKKLESSEENLSSEDEELPCQHLFLGKVNIPQSOTRHSIVATECLSKNTEENL 1260
QY 1261 LSLKNSLDCSNQVILAKASQEHHLSEETKCSASFSSQCELEDLTANTNODPFLIGS 1320
Db 1261 LSLKNSLDCSNQVILAKASQEHHLSEETKCSASFSSQCELEDLTANTNODPFLIGS 1320
QY 1321 SKQMRHQSQGVGLSDKELVSDDEERGTLGLENNOEQSMDSNLGEAASGCESETSVSE 1380
Db 1321 SKQMRHQSQGVGLSDKELVSDDEERGTLGLENNOEQSMDSNLGEAASGCESETSVSE 1380
QY 1381 DCSGLSSQSDILTQORDTMOHNLKLOQMAEAVLEOHGQSPNSYPSIISDSALE 1440
Db 1381 DCSGLSSQSDILTQORDTMOHNLKLOQMAEAVLEOHGQSPNSYPSIISDSALE 1440
QY 1441 DLNRPQSTSEKAVLTQSQSEYPIQONPGLSADKPEVSADSTSKNKEPVGVERSPSK 1500
Db 1441 DLNRPQSTSEKAVLTQSQSEYPIQONPGLSADKPEVSADSTSKNKEPVGVERSPSK 1500
QY 1501 CPSLDRWYMHSCGSLQNNRNPQBELIKVVDVEEQLEESGPHDLTETSYLPQDLEG 1560
Db 1501 CPSLDRWYMHSCGSLQNNRNPQBELIKVVDVEEQLEESGPHDLTETSYLPQDLEG 1560
QY 1561 TPYLESIGLSFDDPSDPSDAPESARVGNIPSTSAKVPQLKVAESAQAHAHT 1620
Db 1561 TPYLESIGLSFDDPSDPSDAPESARVGNIPSTSAKVPQLKVAESAQAHAHT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERYNKRMSMVVSGLTPEERMLVYKFARKKHITLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERYNKRMSMVVSGLTPEERMLVYKFARKKHITLTNLI 1680
QY 1681 TEETHVWMTDAEFVTCERTLKYFLGIAGGKWWVSYFWVTQSIKERKMLNEHDFEVRGDV 1740

Db 1681 TEETHVWMTDAEFVTCERTLKYFLGIAGGKWWVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHOGPKRARESODRKIFRGLIEICCYGPPFTNMPTDQLEWVQLCGASVVKELSSFTL 1800
Db 1741 VNGRNHOGPKRARESODRKIFRGLIEICCYGPPFTNMPTDQLEWVQLCGASVVKELSSFTL 1800
QY 1801 GTGVHPITVVVQPDWATEDNGFHAIGQCEAPVVTREWLDSVALYQCELDYLIQPIPH 1860
Db 1801 GTGVHPITVVVQPDWATEDNGFHAIGQCEAPVVTREWLDSVALYQCELDYLIQPIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 13
US-08-598-591-2
; Sequence 2, Application US/08598591
; Patent No. 5654155
; GENERAL INFORMATION:
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Murphy Patricia D.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: A Consensus Sequence of the Human BRCA1 Gene
; Patent No. 5654155
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince St.
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/598,591
; FILING DATE: herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Swecker, Robert S.
; REGISTRATION NUMBER: 19,885
; REFERENCE/DOCKET NUMBER: 020160-282
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
US-08-598-591-2

Query Match 99.9%; Score 9635; DB 1; Length 1863;
Best Local similarity 99.8%; Pred. No. 0;
Matches 1860; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQVQV INAMQKILCEPICLBELIKEPVSTKCDHIFCKFCMKLNLNKKGPSQ 60
Db 1 MDLSALRVEEVQVQV INAMQKILCEPICLBELIKEPVSTKCDHIFCKFCMKLNLNKKGPSQ 60
QY 61 CPLCKNDITKRSLOESTRESOLVEELKLTICAFQDLDTGLEVANSYNFAKKNNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRESOLVEELKLTICAFQDLDTGLEVANSYNFAKKNNSPEHLKD 120
QY 121 EVSIIQSMGYRNRKRLQSEPNPSLOETSLVSOLSNLGTVRTLTKORTQOPQKTSYVI 180
Db 121 EVSIIQSMGYRNRKRLQSEPNPSLOETSLVSOLSNLGTVRTLTKORTQOPQKTSYVI 180
QY 181 ELGSDSEEDTVNKAATYCSVQDQELLQITPQGTREDEISLDSAKKAACEFSETDVNTNTEHHQ 240
Db 181 ELGSDSEEDTVNKAATYCSVQDQELLQITPQGTREDEISLDSAKKAACEFSETDVNTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPKPYOGSSVSNLHVPCGNTNTHASSLOHENSLLITKDRMVE 300
Db 241 PSNNDLNTTEKRAAERHPKPYOGSSVSNLHVPCGNTNTHASSLOHENSLLITKDRMVE 300
QY 301 KAEFCNKSQOPGLARSOHNRWAGSETCNDRTPTSTKVVLDLADPLCERKWNKQKLPC 360
Db 301 KAEFCNKSQOPGLARSOHNRWAGSETCNDRTPTSTKVVLDLADPLCERKWNKQKLPC 360
QY 361 SENPRDTEVPMTITLNSSIQYNEWFSRDELGLGSDSDHGESESNKAVADVLDVINEVD 420
Db 361 SENPRDTEVPMTITLNSSIQYNEWFSRDELGLGSDSDHGESESNKAVADVLDVINEVD 420
QY 421 EYSGSSEKIDLLASOPHEALICKSERVHSKVESNIEDKIFGKTYRKKASIPNLSHVTEEN 480
Db 421 EYSGSSEKIDLLASOPHEALICKSERVHSKVESNIEDKIFGKTYRKKASIPNLSHVTEEN 480
QY 481 LIIGAEVTEPQIIQERPLTNKLRKRRTSGLHPDEFTKADLAVOKTPEMINOCTNOTE 540
Db 481 LIIGAEVTEPQIIQERPLTNKLRKRRTSGLHPDEFTKADLAVOKTPEMINOCTNOTE 540
QY 541 QNGOVNITNSGHNKTKGDSITONEKNPNP IESLEKESAFKTKABPISSISNMELELNI 600
Db 541 QNGOVNITNSGHNKTKGDSITONEKNPNP IESLEKESAFKTKABPISSISNMELELNI 600
QY 601 HNSKAPKKRLRKSTRIHHALELVVSRNLSPNCTELQIDSCSSSEIEKKKYNQMPV 660
Db 601 HNSKAPKKRLRKSTRIHHALELVVSRNLSPNCTELQIDSCSSSEIEKKKYNQMPV 660
QY 661 RISRNLQMEGKEPATGAKSKNPNQETSKRHSDTTPFELKLTNAPGFTKCSNTSELKE 720
Db 661 RISRNLQMEGKEPATGAKSKNPNQETSKRHSDTTPFELKLTNAPGFTKCSNTSELKE 720
QY 721 FVNPSLPREEKEELETYKVSNNABDPKDLMLSGERVILQTERSVESSISLVPCTDYGTQ 780
Db 721 FVNPSLPREEKEELETYKVSNNABDPKDLMLSGERVILQTERSVESSISLVPCTDYGTQ 780
QY 781 ESISLLEVTILGKATKTEPNKCVSOCAAFENPKGLIHGCKDNRNDTEGFKYPLGHEVNH 840
Db 781 ESISLLEVTILGKATKTEPNKCVSOCAAFENPKGLIHGCKDNRNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLONTFKVSRQSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLONTFKVSRQSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900
QY 901 FCEQKEENQGNESNIPKPVQTNITAGFPVVGQKDKPVDNAKCSIKGSRFCLSSQFRG 960
Db 901 FCEQKEENQGNESNIPKPVQTNITAGFPVVGQKDKPVDNAKCSIKGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPRIPLPFKISFVTKCKKNLLEENFEHNSPSPRENGENIP 1020
Db 961 NETGLITPNKHGLLQNPRIPLPFKISFVTKCKKNLLEENFEHNSPSPRENGENIP 1020
QY 1021 STVSTISRNNIRENVFKKASSNINEVSSSTNEVSSSTNEIGSSDENIQOABLGRNRPKL 1080
Db 1021 STVSTISRNNIRENVFKKASSNINEVSSSTNEVSSSTNEIGSSDENIQOABLGRNRPKL 1080
QY 1081 NAMRLGLVQLOPEVYKQSLPGSNCKHPEIKKOEYEEVQVNTDFSPYLISDNLBQPMGSS 1140

Db 1081 NAMRLGLVQLOPEVYKQSLPGSNCKHPEIKKOEYEEVQVNTDFSPYLISDNLBQPMGSS 1140
QY 1141 HASQVCSFTPDLDLDDGGEIKEDTSPAENDIKESSAVFSKSVQKGLSRSPSFTHTHAQ 1200
Db 1141 HASQVCSFTPDLDLDDGGEIKEDTSPAENDIKESSAVFSKSVQKGLSRSPSFTHTHAQ 1200
QY 1201 GYRRGAKKLESEENLSEDEELPCFQHLFGKVNINISQSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESEENLSEDEELPCFQHLFGKVNINISQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDSCNQVILAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLNDSCNQVILAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHOSQSGVGLSKDELVSDEERGTGLEENNQEQSMDSNLGEAASGESETSVSE 1380
Db 1321 SKQMRHOSQSGVGLSKDELVSDEERGTGLEENNQEQSMDSNLGEAASGESETSVSE 1380
QY 1381 DCSGLSSQSDILTTQORDTMOHNLIKLOEAMAELEAVLEOHGSOQPSNSYPSIISDSSALE 1440
Db 1381 DCSGLSSQSDILTTQORDTMOHNLIKLOEAMAELEAVLEOHGSOQPSNSYPSIISDSSALE 1440
QY 1441 DLRNPEQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSSSTKNKEPGEVERSSPSK 1500
Db 1441 DLRNPEQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSSSTKNKEPGEVERSSPSK 1500
QY 1501 CPSLDDRRWMHSCSGSLQNRNYPQOEELIKVVDVEEQOLESQPHDLTETSYLPRQDLLEG 1560
Db 1501 CPSLDDRRWMHSCSGSLQNRNYPQOEELIKVVDVEEQOLESQPHDLTETSYLPRQDLLEG 1560
QY 1561 TPYLESGLISLSDPESDPEDRAPESARVGNIPSTSTALKVPOLKVAESAQSPAAAHTT 1620
Db 1561 TPYLESGLISLSDPESDPEDRAPESARVGNIPSTSTALKVPOLKVAESAQSPAAAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEBFMLYKFKARKHHTLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEBFMLYKFKARKHHTLTNLI 1680
QY 1681 TEETHVVMKTDABFVCERTLYKFLGTAGGKVVYFVWVTSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDABFVCERTLYKFLGTAGGKVVYFVWVTSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQPKRARSQDRKIFERGLEICCYGPTTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQPKRARSQDRKIFERGLEICCYGPTTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
QY 1801 GTGVHPIVVQPDADWEDNGEFAIGQMCAPVVTREWVLDVALYQCELDYTLIPQIPH 1860
Db 1801 GTGVHPIVVQPDADWEDNGEFAIGQMCAPVVTREWVLDVALYQCELDYTLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863
RESULT 14
US-08-798-691-2
; Sequence 2, Application US/08798691
; Patent No. 5750400
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; TITLE OF INVENTION: BRCA1 Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONCORMED

STREET: 200 Perry Parkway
CITY: Gaithersburg
STATE: MD
COUNTRY: USA
ZIP: 20877
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/798,691
FILING DATE: 12-Feb-97
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Thomas Gallegos
REGISTRATION NUMBER: 32,692
REFERENCE/DOCKET NUMBER: PA-0054CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-527-2051
TELEFAX: 301-208-6997
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCA1
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21

US-08-798-691-2

Query Match 99.9%; Score 9635; DB 1; Length 1863;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1860; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQNVINAMQKILECPICLLELIEPVTCKDHIIFCKFCMLKLLNQKGPQ 60
Db 1 MDLSALRVEEVQNVINAMQKILECPICLLELIEPVTCKDHIIFCKFCMLKLLNQKGPQ 60
QY 61 CPLCKNDITKRSLOESTRFSQVVEELKIIICAFQDITGLYANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVVEELKIIICAFQDITGLYANSYNFAKKENNSPEHLKD 120
QY 121 EYSIIQSMGYRNRKRLQSEPNPSLQETSLSVQLSNLGTVRTLTRKRIQPKTSVYI 180
Db 121 EYSIIQSMGYRNRKRLQSEPNPSLQETSLSVQLSNLGTVRTLTRKRIQPKTSVYI 180
QY 181 ELGSDSSEDTVNKATYCSVGDDQLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHQ 240
Db 181 ELGSDSSEDTVNKATYCSVGDDQLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHQ 240
QY 241 PSNNDLNTTEKRAAERHPKYOGSSVSNLHVPCGNTTHASSLQHENSSILLTKDRMNYE 300
Db 241 PSNNDLNTTEKRAAERHPKYOGSSVSNLHVPCGNTTHASSLQHENSSILLTKDRMNYE 300
QY 301 KAEFCNKSKOPGLARSQHRNAGSKETCNDNRRTPSTPEKKVDNLADPLCKERKNQKLP 360
Db 301 KAEFCNKSKOPGLARSQHRNAGSKETCNDNRRTPSTPEKKVDNLADPLCKERKNQKLP 360
QY 361 SENPRTEDPWITLNSSTQKWNFNRSDDELLGSDSDHGESNAKVADVDLVNEDV 420
Db 361 SENPRTEDPWITLNSSTQKWNFNRSDDELLGSDSDHGESNAKVADVDLVNEDV 420
QY 421 EYSGSEKIDLLASDPHEALICKSERVSHKSVESNTEDEIFGKTYRKKAASLPNLSHVTEN 480
Db 421 EYSGSEKIDLLASDPHEALICKSERVSHKSVESNTEDEIFGKTYRKKAASLPNLSHVTEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPEMINQGTNOTE 540

Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPEMINQGTNOTE 540
QY 541 QNGQVNNITNSGHENKTKGDSIQNEKNPNPIESLKEKSAFKTKAPPISSISNMELELNI 600
Db 541 QNGQVNNITNSGHENKTKGDSIQNEKNPNPIESLKEKSAFKTKAPPISSISNMELELNI 600
QY 601 HNSKAPKNRLRRKSSTRHIIHALELVVSRNLSPNCTELQIDSCSSSEIEIKKKYQNMVP 660
Db 601 HNSKAPKNRLRRKSSTRHIIHALELVVSRNLSPNCTELQIDSCSSSEIEIKKKYQNMVP 660
QY 661 RHSRNQLMEGKEPATGAKKSNKPNEQTSKRHDSDFPELKTNPAGSTKCSNTSELKE 720
Db 661 RHSRNQLMEGKEPATGAKKSNKPNEQTSKRHDSDFPELKTNPAGSTKCSNTSELKE 720
QY 721 FVNPSPREEKEKLETVKVSNNADPKDMLSGERVLQTERSVSSSISLVPGDYGTQ 780
Db 721 FVNPSPREEKEKLETVKVSNNADPKDMLSGERVLQTERSVSSSISLVPGDYGTQ 780
QY 781 ESISLLEVSTLGKAKTEPNKCVSQAAPENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
Db 781 ESISLLEVSTLGKAKTEPNKCVSQAAPENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLQNTFKVSRQSFALFSPGNAEEECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSRQSFALFSPGNAEEECATFSAHSGSLKKQSPKVT 900
QY 901 FECEQKEENQGNESNIKPVQTVNITAGVPVVGOKDPVDNAKCSIKGGRSCLSSQFRG 960
Db 901 FECEQKEENQGNESNIKPVQTVNITAGVPVVGOKDPVDNAKCSIKGGRSCLSSQFRG 960
QY 961 NETGLITPNKHGILLQNPYRIPPLPIKSPVTKCKKNLLEENFEHSMSPEREMGNENIP 1020
Db 961 NETGLITPNKHGILLQNPYRIPPLPIKSPVTKCKKNLLEENFEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNRIENVPFKEASSNINEVGSTNEVGSSINEIGSSDENIQAEIGRNRGPKL 1080
Db 1021 STVSTISRNRIENVPFKEASSNINEVGSTNEVGSSINEIGSSDENIQAEIGRNRGPKL 1080
QY 1081 NAWMLRGVLQPEVYKQSLPGSNCKHPEIKKQVEEVQVQNTDFSPYLLSDNLEQPMGSS 1140
Db 1081 NAWMLRGVLQPEVYKQSLPGSNCKHPEIKKQVEEVQVQNTDFSPYLLSDNLEQPMGSS 1140
QY 1141 HASQVCSETPDDLLDGETIKEDTSAENDIKESSAVFSKQKELSRSPSPFTTHLAQ 1200
Db 1141 HASQVCSETPDDLLDGETIKEDTSAENDIKESSAVFSKQKELSRSPSPFTTHLAQ 1200
QY 1201 GYRRGAKKLESSEENLSSDEELPCFQHLFGKVNIPSQSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSSDEELPCFQHLFGKVNIPSQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSSQCELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSSQCELEDLTANTNTQDPFLIGS 1320
QY 1321 SKOMRHQESQGVGLSDKELVSDDEERGTLGLENQEOGSDSNLGEAASGESETSYS 1380
Db 1321 SKOMRHQESQGVGLSDKELVSDDEERGTLGLENQEOGSDSNLGEAASGESETSYS 1380
QY 1381 DCSGLSSQSDIILTQORDTMQHNLIKLOQEMAELEAVLEBQHGQSQPSNSYPSIISDSSALE 1440
Db 1381 DCSGLSSQSDIILTQORDTMQHNLIKLOQEMAELEAVLEBQHGQSQPSNSYPSIISDSSALE 1440
QY 1441 DLNRNPOSTSEKAVLTQSSEYPIQSONPEGISADKFEVSADSSSTSKNKEPVERSSPSK 1500
Db 1441 DLNRNPOSTSEKAVLTQSSEYPIQSONPEGISADKFEVSADSSSTSKNKEPVERSSPSK 1500
QY 1501 CPSLDDRWMHSCSGSLQNRNYPQOEELIKVVDVBEQOOLEESGPHDLTETSYLPQDLEG 1560
Db 1501 CPSLDDRWMHSCSGSLQNRNYPQOEELIKVVDVBEQOOLEESGPHDLTETSYLPQDLEG 1560
QY 1561 TPYLESIGSLFSDPDSPEDRAPESARVGNIPSSTSALKVPOLKVAESAQSAHAHTT 1620
Db 1561 TPYLESIGSLFSDPDSPEDRAPESARVGNIPSSTSALKVPOLKVAESAQSAHAHTT 1620

Db 1561 TPYLESGISLSDPESDPSEDRAPE SARVGNIPSTSA LKVPOLKVAESAQGPAAHTT 1620
Qy 1621 DTAGNAMEESVSREKPELTASTERVNKRMSVVSGLTPEEFMLVYKFAKKHHTLTNLI 1680
Db 1621 DTAGNAMEESVSREKPELTASTERVNKRMSVVSGLTPEEFMLVYKFAKKHHTLTNLI 1680
Qy 1681 TEETHVVMKTDAEVCERTLKYFLGIAGGKVVSYFWVTQSIKERKMLNHDPEVRGDV 1740
Db 1681 TEETHVVMKTDAEVCERTLKYFLGIAGGKVVSYFWVTQSIKERKMLNHDPEVRGDV 1740
Qy 1741 VNGRHHQPKRARESQRKIFRGLIEICCYGPFTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
Db 1741 VNGRHHQPKRARESQRKIFRGLIEICCYGPFTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
Qy 1801 GTGVPIVVVQPDWTDNGPHATGOMCEAPVTVREXVLDVSVALYQCOELDTYLIPOIPH 1860
Db 1801 GTGVPIVVVQPDWTDNGPHATGOMCEAPVTVREXVLDVSVALYQCOELDTYLIPOIPH 1860
Qy 1861 SHY 1863
Db 1861 SHY 1863

RESULT 15
US-08-798-691-6
; Sequence 6, Application US/08798691
; Patent No. 5750400
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvarez, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; TITLE OF INVENTION: BRCA1 Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONCORMED
; STREET: 200 Ferry Parkway
; CITY: Gaithersburg
; STATE: MD
; COUNTRY: USA
; ZIP: 20877
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,691
; FILING DATE: 12-Feb-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas Gallegos
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: PA-0054CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-527-2051
; TELEFAX: 301-208-6997
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17

MAP POSITION: 17q21
US-08-798-691-6
Query Match 99.9%; Score 9635; DB 1; Length 1863;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1860; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MDLSALRVEEVQNVINAMOKILECPICLELTKEPVSTKCDHIFCKFCMLKLLNKKGPSQ 60
Db 1 MDLSALRVEEVQNVINAMOKILECPICLELTKEPVSTKCDHIFCKFCMLKLLNKKGPSQ 60
Qy 61 CPLCKNDITKRSLQESTRFSQLVVEELLKIIICAFOLDTGLEVANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLQESTRFSQLVVEELLKIIICAFOLDTGLEVANSYNFAKKENNSPEHLKD 120
Qy 121 EVSIQSMGYRNRAKRLQSEPNPSLOETSLVSOLSNLGTVRTLRTKQRTQPKQTSYVI 180
Db 121 EVSIQSMGYRNRAKRLQSEPNPSLOETSLVSOLSNLGTVRTLRTKQRTQPKQTSYVI 180
Qy 181 ELGSDSSEDVTNKATYCSVGDQELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDVTNKATYCSVGDQELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Qy 241 PSNNDLNTTEKRAAERHPEKYQGSVSNLHVPCGTINTHASSLQHENSSLLLTDKRMNVE 300
Db 241 PSNNDLNTTEKRAAERHPEKYQGSVSNLHVPCGTINTHASSLQHENSSLLLTDKRMNVE 300
Qy 301 KAFCNKSKQPGGLARSQHNHWAGSKETCNDRTPTSTKTKVDNLADPLCERKWNKQKLPC 360
Db 301 KAFCNKSKQPGGLARSQHNHWAGSKETCNDRTPTSTKTKVDNLADPLCERKWNKQKLPC 360
Qy 361 SENPRDTEVPWITLNSIOKVNEWFSRDELGLSDSDSHDGESESNKAVADVLDVNLVD 420
Db 361 SENPRDTEVPWITLNSIOKVNEWFSRDELGLSDSDSHDGESESNKAVADVLDVNLVD 420
Qy 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTE 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTE 480
Qy 481 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVQKTPMINOGTNOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVQKTPMINOGTNOTE 540
Qy 541 ONGOVNMTNSGHENKTKGDSIONEKNPNPTIESLEKESAFKTAEPISSSISNMELELNI 600
Db 541 ONGOVNMTNSGHENKTKGDSIONEKNPNPTIESLEKESAFKTAEPISSSISNMELELNI 600
Qy 601 HNSKAPKKNLRRKSSRTHIALELVVSRNLSPPNCTELQIDSCSSSBEIEIKKKYQMPV 660
Db 601 HNSKAPKKNLRRKSSRTHIALELVVSRNLSPPNCTELQIDSCSSSBEIEIKKKYQMPV 660
Qy 661 RHSRNLQMECKEPATGAKKSNKPNEQTSKRHSDTTPPELKLTNAPGSFTKCSNTSELKE 720
Db 661 RHSRNLQMECKEPATGAKKSNKPNEQTSKRHSDTTPPELKLTNAPGSFTKCSNTSELKE 720
Qy 721 FVNPSPREEKEEKLETVKVSNNADPKDMLSGERVLOTERSVESSSISLVPGTDYGTQ 780
Db 721 FVNPSPREEKEEKLETVKVSNNADPKDMLSGERVLOTERSVESSSISLVPGTDYGTQ 780
Qy 781 ESISLLEVTGLGAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNHS 840
Db 781 ESISLLEVTGLGAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNHS 840
Qy 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNPGNAEBECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNPGNAEBECATFSAHSGSLKKQSPKVT 900
Qy 901 FECEQKEENOGKNESNIKPVQTNITAGFPVVGOKDPVDNAKCSIKGGSFCLSSQFRG 960
Db 901 FECEQKEENOGKNESNIKPVQTNITAGFPVVGOKDPVDNAKCSIKGGSFCLSSQFRG 960
Qy 961 NETGLITPNKHGLLQNPYRIPPLFPFKSVKTKCKKNLLENFEEHSMSPEREMGNENIP 1020
Db 961 NETGLITPNKHGLLQNPYRIPPLFPFKSVKTKCKKNLLENFEEHSMSPEREMGNENIP 1020

Db	961	NETGLITPNKHGLLQNPYRIPPLEPIKSFVKTKCKKNLLENFEEHSMSPEREMGNENIP	1020
Qy	1021	STVSTISRNIRENVFKEASSNINEVGSSTINEIGSSDENITQAEIAGRNRPKL	1080
Db	1021	STVSTISRNIRENVFKEASSNINEVGSSTINEIGSSDENITQAEIAGRNRPKL	1080
Qy	1081	NAMLRLGVLOPEVYKQSLPGSNCKHPKIKQOEYEVVQTVNTDFSPYLI	1140
Db	1081	NAMLRLGVLOPEVYKQSLPGSNCKHPKIKQOEYEVVQTVNTDFSPYLI	1140
Qy	1141	HASOVCSSETPDDLDDGCEIREDTSAFENDIKESSAVFSKSVOKGELSRSPFTHHLAQ	1200
Db	1141	HASOVCSSETPDDLDDGCEIREDTSAFENDIKESSAVFSKSVOKGELSRSPFTHHLAQ	1200
Qy	1201	GYRRGAKKLESSEENLSSEDEELPCFOHLLFGKVNINIPQSOTRSTVATECLSKNTEENL	1260
Db	1201	GYRRGAKKLESSEENLSSEDEELPCFOHLLFGKVNINIPQSOTRSTVATECLSKNTEENL	1260
Qy	1261	LSLKNSLNDCSNQVILAKASQEHHLSEBETKCSASLFSQCSSELEDLTANTNTQDPFLIGS	1320
Db	1261	LSLKNSLNDCSNQVILAKASQEHHLSEBETKCSASLFSQCSSELEDLTANTNTQDPFLIGS	1320
Qy	1321	SKQMRHOSQGVGLSDKELVSDDEERGTLGLENNQEQSDMSNLGEAASCCSETSVSE	1380
Db	1321	SKQMRHOSQGVGLSDKELVSDDEERGTLGLENNQEQSDMSNLGEAASCCSETSVSE	1380
Qy	1381	DCSGLSSQSDILTTQQRDTMQHNLIKIQQEMAELEAVLEQHGQSPSNSYPSIISDSSALE	1440
Db	1381	DCSGLSSQSDILTTQQRDTMQHNLIKIQQEMAELEAVLEQHGQSPSNSYPSIISDSSALE	1440
Qy	1441	DLRNPQOSTSEKAVLTQKSSEYPISONPEGLSADKFEVSADSTSKNKEPFGVERSSPSK	1500
Db	1441	DLRNPQOSTSEKAVLTQKSSEYPISONPEGLSADKFEVSADSTSKNKEPFGVERSSPSK	1500
Qy	1501	CPSLDDRWMYHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPRQDLEG	1560
Db	1501	CPSLDDRWMYHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPRQDLEG	1560
Qy	1561	TPYLESIGISLFSDDPESDRAPEARSARVGNIPSTTSALKVPOLKVAESAQAAPAAHTT	1620
Db	1561	TPYLESIGISLFSDDPESDRAPEARSARVGNIPSTTSALKVPOLKVAESAQAAPAAHTT	1620
Qy	1621	DTAGYNAMESVSREKPELFASTERNVKNRMVSVGLTPEEFMLVYKFAKKHHTLTNLI	1680
Db	1621	DTAGYNAMESVSREKPELFASTERNVKNRMVSVGLTPEEFMLVYKFAKKHHTLTNLI	1680
Qy	1681	TEETHVVMKTDAEFVCERTIKYFLGIAGGKWVVSIFYFWVTQSIKERKMLNEHDFEVRGDV	1740
Db	1681	TEETHVVMKTDAEFVCERTIKYFLGIAGGKWVVSIFYFWVTQSIKERKMLNEHDFEVRGDV	1740
Qy	1741	VNGRNHOGPKRARESQRDKIFRGLIEICCYGFTNMPDQLEWMVOLCGASVVKELSSFTL	1800
Db	1741	VNGRNHOGPKRARESQRDKIFRGLIEICCYGFTNMPDQLEWMVOLCGASVVKELSSFTL	1800
Qy	1801	GTGVHPITVVVQPDWEDNGFHAIGQCEAPVVTREWVLDVSVALYQOQELDTYLIPQIPH	1860
Db	1801	GTGVHPITVVVQPDWEDNGFHAIGQCEAPVVTREWVLDVSVALYQOQELDTYLIPQIPH	1860
Qy	1861	SHY	1863
Db	1861	SHY	1863

